The Thesis Committee for Megan Marie Hoklas
Certifies that this is the approved version of the following thesis:

AN INTEGRATED LATENT CONSTRUCT MODELING FRAMEWORK FOR PREDICTING PHYSICAL ACTIVITY ENGAGEMENT AND HEALTH OUTCOMES

APPROVED BY
SUPERVISING COMMITTEE:

Supervisor: ______________________________
Chandra Bhat

____
Randy Machemehl

AN INTEGRATED LATENT CONSTRUCT MODELING FRAMEWORK FOR PREDICTING PHYSICAL ACTIVITY ENGAGEMENT AND HEALTH OUTCOMES

by

Megan Marie Hoklas, B.S.C.E.

Thesis

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Master of Science in Engineering

The University of Texas at Austin

December 2014

Dedication

To my beloved grandparents,

Gloria and Thomas Hoklas

and

Mary and Richard Pomplun

Acknowledgements

I cannot find words to express my gratitude to my advisor, Dr. Chandra Bhat. Without his guidance, mentoring, knowledge and patience this thesis would not have been possible. He inspired me to pursue my graduate degree and has continuously taught me to strive to reach my full potential. I was honored to be mentored by him. I am also grateful to Dr. Randy Machemehl, Dr. Ram Pendyala, Subodh Dubey, Dae Hyun You, and Venu Garikapati, for their shared wisdom and assistance throughout the research process. I would also like to express my gratitude to Lisa Macias for her continual assistance and kindness.

In addition, I was lucky enough to have the absolute support of amazing friends. I am specifically indebted to Christopher Rosas and Melissa Archer, who were always there to offer their love, advice and humor when I needed it most. I am also deeply appreciative for the friendships of Vi Pham, Anh-thu Ho, Alexis Clark, Zeina Wafa, Stasa Zivojnovic, Amy Tran, and Amy Nguyen. They have helped me overcome setbacks and kept me sane throughout this long, but fulfilling road.

Lastly, this journey would not have been achievable without the unconditional love and support of my mother, Sandra Hoklas, my father, Thomas Hoklas, my brother, Matthew Hoklas, and my sister, Michele Hoklas. You have taught me to reach for the moon, and that even if I miss I will land amongst the stars. I cannot express my infinite love and gratitude enough.

Abstract

AN INTEGRATED LATENT CONSTRUCT MODELING FRAMEWORK FOR PREDICTING PHYSICAL ACTIVITY ENGAGEMENT AND HEALTH OUTCOMES

Megan Marie Hoklas, M.S.E.

The University of Texas at Austin, 2014

Supervisor: Chandra Bhat

The health and well-being of individuals is related to their activity-travel patterns. Individuals who undertake physically active episodes such as walking and bicycling are likely to have improved health outcomes compared to individuals with sedentary auto-centric lifestyles. Activity-based travel demand models are able to predict activity-travel patterns of individuals at a high degree of fidelity, thus providing rich information for transportation and public health professionals to infer health outcomes that may be experienced by individuals in various geographic and demographic market segments. However, models of activity-travel demand do not account for the attitudinal factors and lifestyle preferences that affect activity-travel and mode use patterns. Such attitude and preference variables are virtually never collected explicitly in travel surveys, rendering it difficult to include them in model specifications. This paper applies Bhat's (2014) Generalized Heterogeneous Data Model (GHDM) approach, whereby latent constructs representing the degree to which individuals are health conscious and inclined to pursue physical activities may be modeled as a function of observed socio-economic and demographic variables and then included as explanatory factors in models of activity-travel outcomes and walk and bicycle use. The model system is estimated on the 2005-2006 National Health and Nutrition Examination Survey (NHANES) sample, demonstrating the efficacy of the approach and the importance of including such latent constructs in model specifications that purport to forecast activity and time use patterns.

Table of Contents

List of Tables

List of Figures

CHAPTER 1 INTRODUCTION

Recent advances in travel demand modeling have focused on the microsimulation of human activity-travel patterns with a view to better understand how, where, when, and why individuals pursue activities and allocate time to various travel and activity episodes (Arentze and Timmermans, 2008). Models of activity-travel behavior have traditionally used an array of observed explanatory variables to forecast activity-travel and mode usage patterns under a wide variety of scenarios. These include such variables as built environment attributes, network level of service variables, and household and person socio-economic and demographic variables. A missing ingredient in the modeling of human activity-travel patterns continues to be attitudinal constructs that capture the lifestyle preferences and proclivity of individuals (Mokhtarian and Salomon, 2001). In the absence of such constructs, these aspects of human lifestyle are not explicitly accounted for and are presumed to be absorbed in the random error terms of the econometric model chain.

This paper strives to overcome this limitation by presenting a latent variable structural equations modeling approach that includes latent constructs representative of the lifestyle choices or proclivity of individuals. The motivation for this study is derived from the increasing interest in drawing a connection between human activity-travel patterns and public health outcomes. Public health continues to be a top priority globally, with the rise of obesity in both children and adults contributing to studies of active living and how the built environment and the transportation system can be designed to foster active and healthy lifestyles (Frank, 2000). Research has shown that walking and biking are the most popular forms of physical activity and therefore planners are striving to create more walking and biking friendly-communities (Steinman et al, 2010). These strategies are receiving particular attention because individuals who are overweight or obese are at elevated risk of health problems such as coronary heart disease, high blood pressure, stroke, and cancer (National Heart, Lung, and Blood Institute, 2012). According to the U.S. Department of Health and Human Services (1996), the body reacts positively to physical activity leading to reduced risk of adverse medical conditions (Blair et al, 2001). Kimm et al (2005) found that individuals who are less physically active on a regular basis had an increase in BMI (body-mass index), a critical indicator of health. The common recommendation is that individuals should complete at minimum 150 minutes a week of physical activity at a moderate intensity level to promote good health (U.S. Department of Health and Human Services 2008). The goal of these programs and recommendations is to promote healthier lifestyles, physical activity, and good nutrition habits.

Given this interest in the nexus between transportation and public health, it is encouraging to note that activity-based microsimulation models are able to offer rich information about people's activity and travel patterns in time and space at a high degree of fidelity. Specifically, activity-based microsimulation models are able to offer measures of walking, bicycling, and physical (recreational/sport) activity engagement that can be used to assess the potential health implications of an individual's activity-travel and mode usage pattern. However, there may also be unobserved lifestyle preferences and attitudes that impact the amount of walking, bicycling, and physical activity that a person undertakes. Those who are more fitness conscious, love the outdoors, and are active-living oriented are likely to undertake more physical activity than those who are less fitness conscious and more prone to sedentary lifestyles . Travel survey data sets do not include information about such lifestyle preferences and attitudes and hence these variables do not make their way into the activity-based model systems.

This study attempts to address this gap by formulating a latent variable structural equations model system where latent constructs (or variables) representative of an individual's lifestyle are modeled as a function of observed socio-economic and demographic variables routinely collected in travel surveys. These latent constructs, together with socio-economic and demographic variables, are then used to predict an array of activity outcomes that describe how people spend time and pursue activities of various types (including walking, bicycling, and physically active episodes). This is not to say that previous research has not shown that socioeconomic status and attitudinal constructs can influence health outcomes (Gonzalez et al, 2014), but that their results are limited to only include a few indicators such as BMI, age, and exercise in the framework (see also Hendrie et al, 2011). In this research, attitudinal constructs describing health consciousness and physical activity propensity are modeled as latent variables. The model system then utilizes the latent variables as explanatory factors (together with socio-economic and demographic attributes) to predict a series of activity and time use outcomes that are important to public health. The central idea is that activity-based models in practice could potentially be enhanced in the future to include such latent lifestyle constructs, with a view to better predicting how individuals pursue activities, use modes, and spend time for various types of activities.

This is not to say that there is not a vast amount of literature that currently exists at the interface of transportation, specifically active traveling, and public health. This literature identifies not only the influences of demographic and socio-economic characteristics on active travel, physical activity, and health, but that of the built environment and behavioral attitudes as well (Merom et al, 2010; Frank et al, 2007). For example, Lubans et al. (2011) examined past studies documenting the relationship between health-related fitness and active travel to school for children, finding an inverse association between body mass index and active travel to school. Similar pieces found that gender and age impact the child's likelihood to participate in active travel to school, as well as attitudinal perceptions of the child's parents and the child's attitude towards physical activity being other driving factors (Panter et al, 2009). Despite having studies that depict a clear importance between health indicators and physical activity, physical activity and active travel, and health indicators and active travel, this area lacks multi-dimensional dependent outcome models. Current studies have examined the factors associated with changes in body mass index (i.e. age, gender, sedentary behavior such as watching television) (see Basterfield et al, 2010; Mitchell et al, 2013; Zhang et al, 2011) and physical activity (i.e. gender, race, age, sedentary behavior) (see Healy et al, 2008; Bauman et al, 2012). By modeling the dependent outcomes jointly, correlated unobserved factors and any possible casual inter-relationships between endogenous outcomes can be accounted for.

The remainder of this paper is organized as follows. The next section describes the data source used, the data formation process, and the hypothesized impact of the latent variables. The third section details the methodology used in this effort. The fourth section presents estimation results for the latent constructs and the activity outcomes. The fifth and final section offers conclusions and directions for future research.

CHAPTER 2 DATA DESCRIPTION

The data set for this paper is derived from the 2005-2006 National Health and Nutrition Examination Survey (NHANES), which incorporates information from both interviews and physical health examinations into the survey. The NHANES is administered by the National Center for Health Statistics (NCHS) as part of the Centers for Disease Control and Prevention's (CDC) initiative to produce national health statistics. The survey is administered to a nationally representative sample of approximately 5,000 people each year. The 2005-2006 data set, which covers a period of two years, solicited responses from a total of 10,348 individuals. The primary reason for choosing this specific version of the NHANES is that the 2005-2006 survey included a 'Physical Activity and Individual Activities Data' section, which is not included in newer versions of the survey. This section of the survey offered detailed information on specific leisure-time activities pursued by those 12 years of age and older, including data on the specific activity the respondent participated in, the intensity level of the activity, the number of episodes completed within the past month, and average duration of an episode for that activity. The initial dataset downloaded from the NHANES website consisted of responses from a total of 6,408 individuals. After extensive cleaning and filtering of the data, the estimation data set consisted of responses from 4,111 individuals.

Seventeen dependent variables, or outcomes (also sometimes referred to as indicators), related to health or physical activity participation were considered for this study. The first health outcome considered is the blood pressure of the individuals. Systolic and diastolic blood pressure readings were taken for each individual, with repetitive measurements taken to capture the best blood pressure readings. Body mass index is another health outcome used in this study. Trained health technicians measured the body mass index of each participant in kg/m^2 (defined as weight divided by square of the height). Using body mass index groupings provided by the CDC, individuals were placed into one of three categories: underweight and normal weight, overweight, or obese. The next health outcome, self-rated health, refers to how participants of the survey rated their general level of health on a five-point scale (poor, fair, good, very good, and excellent). Another health measure used in this study is an indicator of whether individuals were covered by health insurance or any kind of health care plan. In the final data set used for analysis, 21.2 percent were not covered by any type of insurance or health care plan. Finally, there are three health indicators which specify whether a doctor has ever told the individuals that they have allergies, asthma, or diabetes.

Along with the seven health outcomes, there were ten physical activity participation outcomes included in this study. The physical activities from the survey were classified into four categories: bicycling, walking, moderate "other", and vigorous "other". The "other" category included an assortment of 46 leisure activities (other than walk and bike). Moderate intensity is defined by the survey as causing light sweating or a moderate increase in the heart rate, while vigorous intensity is associated with heavy sweating or a large increase in the heart rate. The first four outcome variables measure the average duration (in minutes per month) of engaging in the four physical activity categories. For modeling purposes, a logarithmic transformation is used for the four continuous duration variables to avoid prediction of negative values. The next four outcome variables, specified as count variables, describe the number of times the individual participated in each of the physical activity groups over the period of a month. The final two indicators considered are the average number of hours that the participants watch television and the number of hours they use the computer in a day. Both indicators had six response categories, i.e., 0 hours, ≤1 hour, >1-2 hours, >2-3 hours, >3-4 hours, and >4 hours.

The explanatory variables included in this study are age, ethnicity, education status, gender, and household annual income. The age and income variables were aggregated into four categories each. The categories for age are: 12-19 years, 20-44 years, 45-64 years, and 65 years or greater, while the categories for income are < $25,000, $25,000-$44,999, $45,000-$64,999, and ≥ $65,000. Occupational information of the respondents is not included as an explanatory variable due to the survey, and the 2000 Census Occupation and Industry Indices, not disclosing which occupations would be classified as being labor intensive physically.

A complete profile of the survey sample used in this study is shown in Table 1. Approximately 17% of the sample respondents completed a bachelor's degree or above, similar to the 19% of the total survey respondents, while the 2006 Census found that 24% of their respondents obtained a bachelor's degree or above (United States Census Bureau, 2006). This difference may be attributed to NHANES oversampling certain subgroups, specifically low-income Caucasians, African-Americans, Hispanics, individuals aged 70 or older, and adolescents aged 12-19, in the survey. It is also found that about one-half of the individuals in the sample have normal blood pressure. However, about two-thirds of the sample are either overweight or obese. More than one-in-five (22 percent) spend four or more hours per day watching television. The average monthly duration in bicycling was about 75 minutes, in walking was about 230 minutes (just shy of four hours), in moderate “other” activities was 387 minutes (about 6.5 hours), and in moderate “vigorous” was 430 minutes (about 7 hours). For the corresponding counts, about 87% of respondents did not have any participation episodes in bicycling (average count of 1.66), 68% did not have any participation episodes in walking (average count of 5.68), 58% did not have any participation episodes in moderate “other” (average count of 7.84), and 61% did not have any participation episodes in vigorous “other” (average count of 8.85).

Two latent variable constructs were developed in this study: Health Consciousness and Physical Activity Propensity. The first latent variable “Health Consciousness” depicts an attitude or awareness towards one’s health. Those who are more health-conscious presumably lead a lifestyle that promotes good health and focus on improving their health. The second latent variable “Physical Activity Propensity” reflects an individual’s natural inclination to participate in physical activities and is representative of the lifestyle (active or not) that an individual adopts. Individuals who have a higher physical activity propensity will tend to participate in more physical activities, and choose to bicycle or walk.

CHAPTER 3 MODEL STRUCTURE

This section presents the model structure and methodology adopted in this paper. First, the modeling framework is described to provide an understanding of how the various indicators and latent constructs are related in an integrated model system. Following the presentation of the framework, the paper details the modeling methodology and formulation.

3.1 MODEL FRAMEWORK

The model framework is presented in Figure 1. There are two sets of exogenous variables. The first set constitute socio-economic and demographic variables such as age, ethnicity, gender, education, and household income. The second set are exogenous health variables such as having asthma, allergies, or diabetes, and having health insurance or a health care plan. While it may be argued that the health conditions (asthma, allergies, and diabetes) are outcomes, it is often the case that these health conditions are not entirely under the control of the individual and may be significantly influenced by family history and heredity (University of Maryland Medical Center, 2011).

There are two health and physical activity related latent constructs that are considered in this paper. Both of these constructs, health consciousness and physical activity propensity, are modeled as a function of socio-economic and demographic variables as well as exogenous health variables. In this particular study, it was found that exogenous health variables were not significant in the physical activity propensity latent variable equation and hence an arrow from exogenous health variables to the physical activity propensity is suppressed in the figure.

The ultimate goal of the equations system is to offer a framework that allows the prediction of the physical activity outcomes, activity and time allocation patterns, and mode usage (walking and bicycling), while explicitly incorporating latent variables or constructs that reflect health consciousness and physical activity propensity (lifestyle). As shown in the figure, the health outcomes are modeled as a function of exogenous health variables, socio-economic and demographic variables, and the two latent constructs of health consciousness and physical activity propensity. Similarly, the ten physical activity outcomes are also modeled as a function of these four entities. A rather notable limitation of this study is the absence of built environment attributes and network level of service characteristics as explanatory factors. The households (respondents) are not geo-coded to any level of geography making it impossible to match such secondary attributes to the records in the data set. On the other hand, this is a unique data set that includes both a series of measures related to health, and a series of variables related to physical activity engagement and bicycling and walking. Reflecting the influence of latent health and lifestyle constructs on health outcomes and physical activity indicators requires a data set that includes these two types of information. For this reason, this data set has been utilized for this research, and future research efforts should aim to include contextual variables in the modeling effort.

The set of health and physical activity outcomes (three health outcomes and ten physical activity outcomes) constitute a mix of dependent variable types including nominal variables, ordered variables, and continuous variables. The modeling methodology needs to accommodate this mixture of variable types in a simultaneous equations modeling framework. In this paper, we use Bhat's (2014) Generalized Heterogeneous Data Model (GHDM) framework for the analysis. The GHDM model proposed in Bhat (2014) accomodates continous, ordinal, and count variables in the measurement equation. However, in addition to these three types of variables, we also have grouped variables. Thus, we add the grouped variable componenet into the measurement equation of GHDM model by reatining the same sets of notations and model progression used by Bhat (2014). There are three components to the proposed model structure: (1) the latent variable structural equations model, and (2) the latent variable measurement equation model, and (3) choice model[1]. Where appropriate, index q for decision-makers is supressed (q=1,2,… Q), and it is assumed that all error terms are independent and identically distributed across decision-makers.

3.2 LATENT VARIABLE STRUCTURAL EQUATIONS MODEL (SEM)

Let l be an index for latent variables (l=1,2,…L). Consider the latent variable z_l^* and write it as a linear function of covariates:

$$z_l^* = \boldsymbol{\alpha}_l' \boldsymbol{w} + \eta_l, \tag{1}$$

where $\boldsymbol{w}$ is a $(\tilde{D} \times 1)$ vector of observed covariates, $\boldsymbol{\alpha_l}$ is a corresponding $(\tilde{D} \times 1)$ vector of coefficients, and η_l is a random error term assumed to be standard normally distributed for identification purposes. Next, define the $(L \times \tilde{D})$ matrix

[1] The structural, measurement equation, and choice model components described here are heavily/completely drawn from Bhat (2014).

$\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \ldots \boldsymbol{\alpha}_L)'$, and the $(L \times 1)$ vectors $\boldsymbol{z}^* = (z_1^*, z_2^*, \ldots, z_L^*)'$ and $\boldsymbol{\eta} = (\eta_1, \eta_2, \eta_3, \ldots, \eta_L)'$. Here, we assume a multivariate normal distribution for the unobserved/error term. That is, $\boldsymbol{\eta} \sim MVN_L[\mathbf{0}_L, \boldsymbol{\Gamma}]$, where $\mathbf{0}_L$ is an $(L \times 1)$ column vector of zeros, and $\boldsymbol{\Gamma}$ is a $(L \times L)$ correlation matrix[2]. In matrix form, equation (1) may be written as:

$$\boldsymbol{z}^* = \alpha w + \boldsymbol{\eta} \tag{2}$$

3.3 LATENT VARIABLE MEASUREMENT EQUATION MODEL COMPONENTS

The measurement equation of the GHDM is a system of simultaneous equations as it consist of four types of variables (continous, ordinal, grouped, and count). Thus, we derive the expression for each variable seperately in turn and then finally join them using matrix notations (see section 3.4).

3.3.1 Continous Variable Component

Let there be H continuous outcomes $(y_1, y_2, \ldots, y_H)$ with an associated index h $(h = 1, 2, \ldots, H)$. Then, in the usual linear regression fashion, we can write:

$$y_h = \boldsymbol{\gamma}_h' \boldsymbol{x} + \boldsymbol{d}_h' \boldsymbol{z}^* + \varepsilon_h \tag{3}$$

---

[2] There are two ways to fix the scale of the latent variable: (1) by fixing the diagonal elements of error matrix to unity (i.e., correlation matrix) and estimate all the coefficients corresponding to loading of latent variables on the outcomes/indicators in the measurement equation, and (2) by fixing one of the coefficient of latent variable loading on the indicator/outcome to be unity for each of the latent variable in the measurement equation and estimate a unrestricted error covariance matrix (see, Stapleton, 1978). In this study, we use the former way (i.e., a correlation error matrix) for fixing the scale. Methodologically, there should not be any difference in terms of results (i.e., estimated parameters and their directions) due to former or later way of ensuring identification. However, Raveau et al., (2012) observed that constraining the measurement equation parameters can cause serious problems as compared to constraining error covariance elements such as wrong parameter direction, etc.

where $\boldsymbol{x}$ is a $(A \times 1)$ vector of exogenous variables (including a constant) as well as possible endogenous variables, $\boldsymbol{\gamma}_h$ is the corresponding column vector of coefficients, $\boldsymbol{d}_h$ is an $(L \times 1)$ vector of latent variable loadings on the h^{th} continuous outcome, and ε_h is a normally distributed error term. Stack the H continuous outcomes into an $(H \times 1)$ vector $\boldsymbol{y}$, and the H error terms into another $(H \times 1)$ vector $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_H)'$. Also, let $\boldsymbol{\Sigma}$ be the covariance matrix of $\boldsymbol{\varepsilon}$, which is restricted to be diagonal (it is due to the identification issues. See, Bhat (2014) for a detailed discussion on various identification issues or Reilly and O'Brien (1996) for a detailed discussion on identification issues related to structural equation modeling). Define the $(H \times A)$ matrix $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \ldots \gamma_H)'$ and the $(H \times L)$ matrix of latent variable loadings $\boldsymbol{d} = (\boldsymbol{d_1}, \boldsymbol{d_2}, \ldots, \boldsymbol{d_H})'$. Then, in matrix form, the continous variable component can be written as follows:

$$\boldsymbol{y} = \boldsymbol{\gamma x} + \boldsymbol{dz}^* + \boldsymbol{\varepsilon}. \quad (4)$$

3.3.2 Ordinal Variable Component

Let there be N ordinal outcomes with an associated index n $(n = 1, 2, \ldots, N)$. Also, let J_n be the number of categories for the n^{th} ordinal outcome $(J_n \geq 2)$ and let the corresponding index be $j_n (j_n = 1, 2, \ldots, J_n)$. Let $\tilde{y}_n^*$ be the latent underlying variable whose horizontal partitioning leads to the observed category for the n^{th} ordinal variable. Assume that the individual under consideration chooses the a_n^{th} ordinal category. Then, in the usual ordered response formulation, one can write:

$$\tilde{y}_n^* = \tilde{\boldsymbol{\gamma}}_n' \boldsymbol{x} + \tilde{\boldsymbol{d}}_n' \boldsymbol{z}^* + \tilde{\varepsilon}_n, \text{ and } \tilde{\psi}_{n,a_n-1} < \tilde{y}_n^* < \tilde{\psi}_{n,a_n}, \tag{5}$$

where $\boldsymbol{x}$ is a $(A \times 1)$ vector of exogenous variables and possible endogenous variables, $\tilde{\boldsymbol{\gamma}}_n$ is the corresponding column vector of coefficients, $\tilde{\boldsymbol{d}}_n$ is an $(L \times 1)$ vector of latent variable loadings on the n^{th} ordinal outcome, the $\tilde{\psi}$ terms represent thresholds, a_n is the observed ordinal variable category, and $\tilde{\varepsilon}_n$ is the standard normal random error term. Further, the threshold needs to be in increasing order as follows: $\tilde{\psi}_{n,0} < \tilde{\psi}_{n,1} < \tilde{\psi}_{n,2} \ldots < \tilde{\psi}_{n,J_n-1} < \tilde{\psi}_{n,J_n}$; $\tilde{\psi}_{n,0} = -\infty, \tilde{\psi}_{n,1} = 0$, and $\tilde{\psi}_{n,J_n} = +\infty$. For later use, let $\tilde{\boldsymbol{\psi}}_n = (\tilde{\psi}_{n,2}, \tilde{\psi}_{n,3} \ldots, \tilde{\psi}_{n,J_n-1})'$ *and* $\tilde{\boldsymbol{\psi}} = (\tilde{\boldsymbol{\psi}}_1', \tilde{\boldsymbol{\psi}}_2' \ldots, \tilde{\boldsymbol{\psi}}_N)'$. Stack the N underlying continous variables $\tilde{y}_n^*$ into an $(N \times 1)$ vector $\tilde{\boldsymbol{y}}^*$, and the N error terms $\tilde{\varepsilon}_n$ into another $(N \times 1)$ vector $\tilde{\boldsymbol{\varepsilon}}$. Define $\tilde{\boldsymbol{\gamma}} = (\tilde{\gamma}_1, \tilde{\gamma}_2 \ldots, \tilde{\gamma}_N)'$ [$(N \times A)$ matrix] and $\tilde{\boldsymbol{d}} = (\tilde{\boldsymbol{d}}_1, \tilde{\boldsymbol{d}}_2, \ldots, \tilde{\boldsymbol{d}}_N)$ [$(N \times L)$ matrix], and let $\mathbf{IDEN}_N$ be the identity matrix of dimension N representing the correlation matrix of $\tilde{\boldsymbol{\varepsilon}}$ (so, $\tilde{\boldsymbol{\varepsilon}} \sim MVN_N(\mathbf{0}_N, \mathbf{IDEN}_N)$). Finally, stack the lower thresholds $\tilde{\psi}_{n,a_n-1}(n = 1, 2, \ldots, N)$ into an $(N \times 1)$ vector $\tilde{\boldsymbol{\psi}}_{low}$ and the upper thresholds $\tilde{\psi}_{n,a_n}(n = 1, 2, \ldots, N)$ into another vector $\tilde{\psi}_{up}$. Then, in matrix form, the ordinal variable component can be written as follows:

$$\tilde{\boldsymbol{y}}^* = \tilde{\boldsymbol{\gamma}} \boldsymbol{x} + \tilde{\boldsymbol{d}} \boldsymbol{z}^* + \tilde{\boldsymbol{\varepsilon}}, \quad \tilde{\boldsymbol{\psi}}_{low} < \tilde{\boldsymbol{y}}^* < \tilde{\boldsymbol{\psi}}_{up}. \tag{6}$$

3.3.3 Grouped Variable Component

The grouped variable is similar to the ordinal variable, except in this case, the thresholds are known and need not be estimated. Instead, to fix the scale of the variable, the variance of the grouped variable is estimated (see, Bhat; 1994). Let there be V grouped outcomes with an associated index v $(v = 1, 2, ..., V)$. Also, let J_v be the number of categories for the v^{th} grouped outcome $(J_v \geq 2)$ and let the corresponding index be $j_v(j_v = 1, 2, ..., J_v)$. Let $\vec{y}_v^*$ be the latent underlying variable whose horizontal partitioning based on the known thresholds leads to the observed category for the v^{th} grouped variable. Assume that the analyst specifices/observes the a_v^{th} category for the v^{th} grouped variable based on the known thresholds. Then, in the usual ordered response formulation, one can write:

$$\vec{y}_v^* = \vec{\boldsymbol{\gamma}}_v' \boldsymbol{x} + \vec{\boldsymbol{d}}_v' \boldsymbol{z}^* + \vec{\varepsilon}_v, \; and \;\; \vec{\psi}_{v,a_v-1} < \vec{y}_v^* < \vec{\psi}_{v,a_v}, \tag{7}$$

where $\boldsymbol{x}$ is a $(A \times 1)$ vector of exogenous variables and possible endogenous variables, $\vec{\boldsymbol{\gamma}}_v$ is the corresponding column vector of coefficients, $\vec{\boldsymbol{d}}_v$ is an $(L \times 1)$ vector of latent variable loadings on the v^{th} grouped outcome, the $\vec{\psi}$ terms represent thresholds, and $\vec{\varepsilon}_v$ is the standard normal random error term. Similar, to the ordinal variable component, the thresholds needs to be in increasing order. Stack the thresholds for each grouped variable in a vector $\vec{\boldsymbol{\psi}}_v$ as follow: $\vec{\boldsymbol{\psi}}_v = (\vec{\psi}_{v,2}, \vec{\psi}_{v,3} ..., \vec{\psi}_{v,J_v-1})'$. Stack the V underlying continuous variables $\vec{y}_v^*$ into an $(V \times 1)$ vector $\vec{\boldsymbol{y}}^*$, and the V error terms $\vec{\varepsilon}_v$ into another $(V \times 1)$ vector $\vec{\boldsymbol{\varepsilon}}$. Define $\vec{\boldsymbol{\gamma}} = (\vec{\boldsymbol{\gamma}}_1, \vec{\boldsymbol{\gamma}}_2 ..., \vec{\boldsymbol{\gamma}}_V)'$ [$(V \times A)$ matrix] and $\vec{\boldsymbol{d}} = \left(\vec{\boldsymbol{d}}_1, \vec{\boldsymbol{d}}_2, ..., \vec{\boldsymbol{d}}_V\right)$ [$(V \times L)$ matrix], and let $\boldsymbol{\Xi}$ be the diagonal matrix of dimension V representing the covariance matrix of $\vec{\boldsymbol{\varepsilon}}$. That is,

$$\boldsymbol{\Xi} = \begin{bmatrix} \sigma_1^2 & 0 & 0 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & 0 & 0 & \cdots & 0 \\ 0 & 0 & \sigma_3^2 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ldots & \vdots \\ 0 & 0 & 0 & 0 & \ldots & \sigma_V^2 \end{bmatrix} \tag{8}$$

Stack the lower thresholds for each grouped outcome $\vec{\psi}_{v,a_v-1}(v = 1, 2, \ldots, V)$ into an $(V \times 1)$ vector $\vec{\boldsymbol{\psi}}_{\boldsymbol{low}}$ and the upper thresholds $\vec{\psi}_{v,a_v}(v = 1, 2, \ldots, V)$ into another vector $\vec{\psi}_{up}$. Then, in matrix form, the grouped variable component can be written as follows:

$$\vec{\boldsymbol{y}}^* = \vec{\boldsymbol{\gamma}}\boldsymbol{x} + \vec{\boldsymbol{d}}\boldsymbol{z}^* + \vec{\boldsymbol{\varepsilon}}, \;\; \vec{\boldsymbol{\psi}}_{\boldsymbol{low}} < \vec{\boldsymbol{y}}^* < \vec{\boldsymbol{\psi}}_{\boldsymbol{up}} \tag{9}$$

3.3.4 Count Variable Component

Let there be C count variables with an associated index c $(c = 1, 2, \ldots, C)$. Let the count index be k_c $(k_c = 0, 1, 2, \ldots, \infty)$ and let r_c be the actual observed count value. Then, a generalized version of the negative binomial count model in a generalized ordered-response probit formulation (Castro et al, 2011; Bhat et al, 2013), may be written as:

$$\breve{y}_c^* = \breve{\boldsymbol{d}}_{\boldsymbol{c}}'\boldsymbol{z}^* + \breve{\varepsilon}_c, \; \breve{\psi}_{c,r_c-1} < \breve{y}_c^* < \breve{\psi}_{c,r_c}, \tag{10}$$

$$\breve{\psi}_{c,r_c} = \Phi^{-1}\left[\frac{(1-\upsilon_c)^{\theta_l}}{\Gamma(\theta_c)} \sum_{t=0}^{r_c} \left(\frac{\Gamma(\theta_c + t)}{t!} (\upsilon_c)^t \right) \right] + \varphi_{c,r_c}, \; \upsilon_c = \frac{\lambda_c}{\lambda_c + \theta_c}, \text{ and } \boldsymbol{\lambda}_c = e^{\breve{\gamma}_c'\boldsymbol{x}} \tag{11}$$

where $\breve{y}_c^*$ is a latent continuous stochastic propensity variable associated with the count variable c that maps into the observed count r_c through the $\breve{\boldsymbol{\psi}}_{\boldsymbol{c}}$ vector (which is a vertically stacked column vector of thresholds $(\breve{\psi}_{c,-1}, \breve{\psi}_{c,0}, \breve{\psi}_{c,1}, \breve{\psi}_{c,2}, \ldots)'$). $\breve{\boldsymbol{d}}_{\boldsymbol{c}}$ is an $(L \times 1)$

vector of latent variable loadings on the c^{th} count outcome, and $\breve{\varepsilon}_c$ is a standard normal random error term. $\boldsymbol{x}$ is a $(A \times 1)$ vector of exogenous variables and possible endogenous variables, $\breve{\boldsymbol{\gamma}}_c$ is the correponding column vector of coefficients. Φ^{-1} in the threshold function of Equation (11) is the inverse function of the univariate cumulative standard normal. θ_c is a parameter that provides flexibility to the count formulation, and is related to the dispersion parameter in a traditional negative binomial model $(\theta_c > 0 \forall c)$. $\Gamma(\theta_c)$ is the traditional gamma function; $\Gamma(\theta_c) = \int_{\tilde{t}=0}^{\infty} \tilde{t}^{\theta_c - 1} e^{-\tilde{t}} \, d\tilde{t}$. As usual, thresholds needs to be in increasing order, which can be ensured by maintaing φ elements in increasing order (i.e., $\varphi_{c,-1} < \varphi_{c,0} < \varphi_{c,1} < \varphi_{c,2} .... < \infty.$) for each count outcome. For identification, we set $\varphi_{c,-1} = -\infty$ and $\varphi_{c,0} = 0$ for all count variables c. In addition, we identify a count value e_c^* $(e_c^* \in \{0,1,2,......\})$ above which φ_{c,k_c} $(k_c \in \{1,2,......\})$ is held fixed at φ_{k,e_c^*}; that is, $\varphi_{c,k_c} = \varphi_{c,e_c^*}$ if $k_c > e_c^*$, where the value of e_c^* can be based on empirical testing. For later use, let $\boldsymbol{\varphi}_c = (\varphi_{c,1}, \varphi_{c,2}, \ldots \varphi_{c,e_c^*})'$ $(e_c^* \times 1$ vector) (assuming $e_c^* > 0)$, $\boldsymbol{\varphi} = (\boldsymbol{\varphi}_1', \boldsymbol{\varphi}_2', \ldots \boldsymbol{\varphi}_C')'$ $\left[\left(\sum_c e_c^* \right) \times 1 \text{ vector} \right]$, and $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots \theta_C)'$ $[\text{C} \times 1 \text{ vector}]$. Also, stack the C latent variables $\breve{y}_c^*$ into a $(C \times 1)$ vector $\breve{\boldsymbol{y}}^*$, and the C error terms $\breve{\varepsilon}_c$ into another $(C \times 1)$ vector $\breve{\boldsymbol{\varepsilon}}$. Let $\breve{\boldsymbol{\varepsilon}} \sim MVN_C(\mathbf{0}_C, \mathbf{IDEN}_\mathbf{C})$, and stack the lower thresholds of the individual $\breve{\psi}_{c,r_c - 1} (c = 1, 2, ..., C)$ into a $(C \times 1)$ vector $\breve{\boldsymbol{\psi}}_{low}$, and the upper thresholds

$\breve{\psi}_{c,r_c}\,(c = 1, 2, \ldots, C)$ into another $(C \times 1)$ vector $\breve{\boldsymbol{\psi}}_{\boldsymbol{up}}$. Define $\breve{\boldsymbol{\gamma}} = (\breve{\gamma}_1, \breve{\gamma}_2, \ldots, \breve{\gamma}_C)'$ [$(C \times A)$ matrix] and $\breve{\boldsymbol{d}} = \left(\breve{\boldsymbol{d}}_{\boldsymbol{1}}, \breve{\boldsymbol{d}}_{\boldsymbol{2}}, \ldots, \breve{\boldsymbol{d}}_{\boldsymbol{C}}\right)'$ [$(C \times L)$ matrix]. Then, in matrix form, the count variable component can be written as follows:

$$\breve{\boldsymbol{y}}^* = \breve{\boldsymbol{d}}\boldsymbol{z}^* + \breve{\boldsymbol{\varepsilon}}, \quad \breve{\boldsymbol{\psi}}_{\boldsymbol{low}} < \boldsymbol{h}^* < \breve{\boldsymbol{\psi}}_{\boldsymbol{up}} \tag{12}$$

3.3.5 Choice Model

Let there be G nominal (unordered-response) variables with an associated index g (g = 1, 2, 3, …, G). Also, let I_g be the number of alternatives corresponding to the g^{th} nominal variable ($I_g \geq 2$) and let i_g be the corresponding index (i_g = 1, 2, 3, …, I_g). Consider the g^{th} nominal variable and assume that the individual chooses the alternative m_g. Then, in usual random utility structure, the utility equation for each alternative i_g can be writen as :

$$\boldsymbol{U}_{\boldsymbol{gi_g}} = \boldsymbol{b}'_{gi_g}\boldsymbol{x} + \vartheta'_{gi_g}(\boldsymbol{\beta}_{gi_g}\boldsymbol{z}^*) + \varsigma_{gi_g}, \tag{13}$$

where $\boldsymbol{x}$ is a $(A \times 1)$ vector of exogenous variables and possible endogenous variables, $\boldsymbol{b}_{gi_g}$ is the correponding column vector of coefficients, and ς_{gi_g} is a normal error term. $\boldsymbol{\beta}_{gi_g}$ is an $(N_{gi_g} \times L)$-matrix of variables interacting with latent variables to influence the utility of alternative i_g, and $\vartheta_{\boldsymbol{gi_g}}$ is an $(N_{gi_g} \times 1)$-column vector of coefficients capturing the effects of latent variables and their interaction effects with other exogenous variables. Let $\boldsymbol{\varsigma_g} = (\varsigma_{g1}, \varsigma_{g2}, \ldots \varsigma_{gI_g})'$ ($I_g \times 1$ vector), and $\boldsymbol{\varsigma_g} \sim MVN_{I_g}(0, \boldsymbol{\Lambda_g})$. Since only the difference in utility matters, only the elements of differenced error covariance matrix $\breve{\Lambda}_g$ are estimatble. That is, $\breve{\boldsymbol{\varsigma}}_{\boldsymbol{g}} = (\breve{\varsigma}_{g2}, \breve{\varsigma}_{g3}, ..., \breve{\varsigma}_{gI_g})$ (where $\breve{\varsigma}_{gi} = \breve{\varsigma}_{gi} - \breve{\varsigma}_{g1}$, $\boldsymbol{i} \neq \mathbf{1}$). Further, the variance term at the top left diagonal of $\breve{\Lambda}_g$ (g=1,2,….,G) is set to 1 to account for scale invariance. However, the error differenced matrix $\breve{\boldsymbol{\Lambda}}_{\mathbf{g}}$ should be constructed from un-differenced error matrix $\boldsymbol{\Lambda}_{\mathbf{g}}$. To do that, add a row of zeros on top and a column of zeros to the left of error differenced matrix $\breve{\boldsymbol{\Lambda}}_{\mathbf{g}}$. In addition, one of the alternatives serves as the base when introducing alternative-specific constants and variables that do not vary across alternatives (that is, whenever an element of $\boldsymbol{x}$ is individual-specific and not alternative-specific, the corresponding element in $\boldsymbol{b}_{gi_g}$ is set to zero for at least one alternative i_g). To proceed, define $\boldsymbol{U}_g = (U_{g1}, U_{g2}, ..., U_{gI_g})'$ ($I_g \times 1$ vector), $\boldsymbol{b}_g = (\boldsymbol{b}_{\boldsymbol{g}1}, \boldsymbol{b}_{\boldsymbol{g}2}, \boldsymbol{b}_{\boldsymbol{g}3}, ..., \boldsymbol{b}_{\boldsymbol{g}I_g})'$ ($I_g \times A$ matrix), and $\boldsymbol{\beta}_g = (\boldsymbol{\beta}'_{g1}, \boldsymbol{\beta}'_{g2}, ..., \boldsymbol{\beta}'_{gI_g})'$ $\left(\sum_{i_g=1}^{I_g} N_{gi_g} \times L\right)$ matrix. Also, define the $\left(I_g \times \sum_{i_g=1}^{I_g} N_{gi_g}\right)$ matrix $\vartheta_{\boldsymbol{g}}$, which is initially filled with all zero values. Then, position the $(1 \times N_{g1})$ row vector $\vartheta'_{\boldsymbol{g1}}$ in the first row to occupy columns 1 to N_{g1}, position the $(1 \times N_{g2})$ row vector $\vartheta'_{\boldsymbol{g2}}$ in the second row to occupy columns $N_{g1}+1$ to $N_{g1} + N_{g2}$, and so on until the $(1 \times N_{gI_g})$ row vector $\vartheta'_{\boldsymbol{gI_g}}$ is appropriately positioned. Further, define $\varpi_g = (\vartheta_g \boldsymbol{\beta}_g)$ ($I_g \times L$ matrix), $\ddot{G} = \sum_{g=1}^{G} I_g$, $\tilde{G} = \sum_{g=1}^{G} (I_g - 1)$, $\tilde{T} = \sum_{g=1}^{G} T_g$,

$\boldsymbol{U} = (\boldsymbol{U}_1', \boldsymbol{U}_2', \ldots, \boldsymbol{U}_G')'$ $(\vec{G} \times 1$ vector), $\varsigma = (\varsigma_1, \varsigma_2, \ldots \varsigma_G)'$ $(\vec{G} \times 1$ vector), $\boldsymbol{b} = (\boldsymbol{b}_1', \boldsymbol{b}_2', \ldots, \boldsymbol{b}_G')'$ $(\vec{G} \times A$ matrix), $\varpi = (\varpi_1', \varpi_2', \ldots, \varpi_G')'$ $(\vec{G} \times L$ matrix), and $\vartheta_{vec} = Vech(\vartheta_1, \vartheta_2, \ldots, \vartheta_G)$ (that is, ϑ_{vec} is a column vector that includes all elements of the matrices $\vartheta_1, \vartheta_2, \ldots, \vartheta_G$). Then, in matrix form, the Equation (13) can be written as follows:

$$\boldsymbol{U} = \boldsymbol{bx} + \varpi \boldsymbol{z}^* + \varsigma, \tag{14}$$

where $\varsigma \sim MVN_{\vec{G}}(0_{\vec{G}}, \boldsymbol{\Lambda})$. As earlier, to ensure identification, we specify $\boldsymbol{\Lambda}$ as follows:

$$\boldsymbol{\Lambda} = \begin{bmatrix} \boldsymbol{\Lambda}_1 & 0 & 0 & 0. \cdots & 0 \\ 0 & \boldsymbol{\Lambda}_2 & 0 & 0 \cdots & 0 \\ 0 & 0 & \boldsymbol{\Lambda}_3 & 0 \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \ldots & \vdots \\ 0 & 0 & 0 & 0 \ldots & \boldsymbol{\Lambda}_G \end{bmatrix} (\vec{G} \times \vec{G} \text{ matrix}), \tag{15}$$

In the general case, this allows the estimation of $\sum_{g=1}^{G} \left(\frac{I_g * (I_g - 1)}{2} - 1 \right)$ terms across all the G nominal variables, as originating from $\left(\frac{I_g * (I_g - 1)}{2} - 1 \right)$ terms embedded in each $\breve{\boldsymbol{\Lambda}}_g$ matrix; g=1,2,…G)[3] .

3.4 Model System Identification and Estimation

Let $E = (H + N + V + C)$. Define $\bar{\boldsymbol{y}} = \left(\boldsymbol{y}', [\tilde{\boldsymbol{y}}^*]', [\vec{\boldsymbol{y}}^*]', [\breve{\boldsymbol{y}}^*]' \right)'$ $[E \times 1 \; vector]$, $\bar{\boldsymbol{\gamma}} = (\boldsymbol{\gamma}', \tilde{\boldsymbol{\gamma}}', \vec{\boldsymbol{\gamma}}', \boldsymbol{0}_{AC})'$ $[E \times A$ matrix], $\bar{\boldsymbol{d}} = (\boldsymbol{d}', \tilde{\boldsymbol{d}}', \vec{\boldsymbol{d}}', \breve{\boldsymbol{d}}')'$ $[E \times L$ matrix], and

[3] If all the un-ordered outcomes are binary variables, then there are no elements to be estimated in the matrix $\boldsymbol{\Lambda}$.

$\bar{\boldsymbol{\varepsilon}} = (\boldsymbol{\varepsilon}', \tilde{\boldsymbol{\varepsilon}}', \vec{\boldsymbol{\varepsilon}}', \breve{\boldsymbol{\varepsilon}}')'$ ($E \times 1$ vector) where $\boldsymbol{0}_{AC}$ is a matrix of zeros of dimension $A \times C$. Let $\boldsymbol{\delta}$ be the collection of parameters to be estimated: $\boldsymbol{\delta} = [\text{Vech}(\boldsymbol{\alpha}), \text{Vech}(\Sigma), \text{Vech}(\vec{\gamma}), \text{Vech}(\vec{\boldsymbol{d}}), \boldsymbol{\psi}, \text{Vech}(\breve{\gamma}), \boldsymbol{\phi}, \boldsymbol{\theta}, \text{Vech}(\boldsymbol{b}), \text{Vech}(\varpi), \text{Vech}(\boldsymbol{\Lambda})]$, where the operator "Vech(.)" vectorizes all the non-zero elements of the matrix/vector on which it operates. With the help of these definitions, the individual components of the GHDM can be written compactly as:

$$\boldsymbol{z}^* = \boldsymbol{\alpha w} + \boldsymbol{\eta} \tag{15}$$

$$\bar{\boldsymbol{y}} = \bar{\boldsymbol{\gamma}}\boldsymbol{x} + \bar{\boldsymbol{d}}\boldsymbol{z}^* + \bar{\boldsymbol{\varepsilon}}, \text{ with } \text{Var}(\bar{\boldsymbol{\varepsilon}}) = \bar{\boldsymbol{\Sigma}} = \begin{bmatrix} \boldsymbol{\Sigma} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \textbf{IDEN}_{\textbf{N}} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{\Xi} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \textbf{IDEN}_{\textbf{C}} \end{bmatrix} (E \times E \textit{ matrix}) \tag{16}$$

$$\boldsymbol{U} = \boldsymbol{bx} + \boldsymbol{\varpi z}^* + \boldsymbol{\varsigma} \tag{17}$$

To develop the reduced form equations, substitute the value of $\boldsymbol{z}^*$ from Equation (15) into Equation (16) and (17).

$$\bar{\boldsymbol{y}} = \bar{\boldsymbol{\gamma}}\boldsymbol{x} + \bar{\boldsymbol{d}}\boldsymbol{z}^* + \bar{\boldsymbol{\varepsilon}} = \bar{\boldsymbol{\gamma}}\boldsymbol{x} + \bar{\boldsymbol{d}}(\boldsymbol{\alpha w} + \boldsymbol{\eta}) + \bar{\boldsymbol{\varepsilon}} = \bar{\boldsymbol{\gamma}}\boldsymbol{x} + \bar{\boldsymbol{d}}\boldsymbol{\alpha w} + \bar{\boldsymbol{d}}\boldsymbol{\eta} + \bar{\boldsymbol{\varepsilon}} \tag{18}$$

$$\boldsymbol{U} = \boldsymbol{bx} + \boldsymbol{\varpi z}^* + \boldsymbol{\varsigma} = \boldsymbol{bx} + \boldsymbol{\varpi}(\boldsymbol{\alpha w} + \boldsymbol{\eta}) + \boldsymbol{\varsigma} = \boldsymbol{bx} + \boldsymbol{\varpi\alpha w} + \boldsymbol{\varpi\eta} + \boldsymbol{\varsigma}$$

Define $\boldsymbol{yU} = [\bar{\boldsymbol{y}}', \boldsymbol{U}']'$ a $[(E + \vec{G}) \times 1)]$ vector. Then $\boldsymbol{yU} \sim \boldsymbol{MVN}_{E+\ddot{G}}(\boldsymbol{B}, \boldsymbol{\Omega})$.

$$\text{where } \boldsymbol{B} = \begin{bmatrix} \boldsymbol{B}_1 \\ \boldsymbol{B}_2 \end{bmatrix} = \begin{bmatrix} \vec{\boldsymbol{\gamma}}\boldsymbol{x} + \vec{\boldsymbol{d}}\boldsymbol{\alpha w} \\ \boldsymbol{bx} + \boldsymbol{\varpi}\,\boldsymbol{\alpha w} \end{bmatrix} \text{ and } \boldsymbol{\Omega} = \begin{bmatrix} \bar{\boldsymbol{d}}\boldsymbol{\Gamma}\bar{\boldsymbol{d}}' + \bar{\boldsymbol{\Sigma}} & \bar{\boldsymbol{d}}\boldsymbol{\Gamma}\boldsymbol{\varpi}' \\ \boldsymbol{\varpi}\boldsymbol{\Gamma}\bar{\boldsymbol{d}}' & \boldsymbol{\varpi}\boldsymbol{\Gamma}\boldsymbol{\varpi}' + \boldsymbol{\Lambda} \end{bmatrix} \tag{19}$$

Equation (19) is now exactly in the same form as provided in Bhat (2014) and can be solve using the estimation approach shown in Bhat (2014). We do not present the estimation approach here, as it only leads to repition. However, the estimation approach as described in Bhat (2014) is provided in Appendix A for readers convenience. Further, readers are referred to Bhat (2014) for a detailed discussion on identification issues.

CHAPTER 4 MODEL ESTIMATION RESULTS

This section presents model estimation results. Results are presented first for the latent variable structural equation estimation, followed by the results for latent variable measurement equation components (non-nominal and nominal outcomes).

4.1 Endogenous Effects

These effects correspond to recursive effects among the endogenous outcomes . These are parts of the $\boldsymbol{\gamma}$ matrix (for the continuous variables), $\tilde{\boldsymbol{\gamma}}$ matrix (for the ordinal variables), $\vec{\boldsymbol{\gamma}}$ matrix (for the grouped variables), the $\breve{\boldsymbol{\gamma}}$ matrix (for the count variables), and the $\boldsymbol{b}$ matrix (for the nominal variables), and represent "cleansed" effects after accommodating unobserved covariance effects through the latent variables discussed in the previous section. The final directions of the recursive effects are obtained after an extensive testing of various model specifications, and choosing the specification that provided the best data fit in terms of the composite marginal log-likelihood value (note, however, that regardless of the presence or absence of recursive effects, the model is a joint model because of the presence of latent variables that impact the many dependent variables). Figure 2 presents the directions of the endogenous relationships. Our results indicate that, after accommodating the jointness among the dependent variables caused by the latent variables, the television and computer duration affects body mass index (BMI), and both of these attributes in turn impacts duration and episodes of walking, moderate physical activity, and vigorous physical activity. Finally, the self-reported health and blood pressure of an individual is determined using aforementioned variables.

4.2 LATENT VARIABLE STRUCTURAL EQUATION MODEL RESULTS

Table 2 provides the results for the effects of individual-specific variables on the two latent constructs in the structural equation model. In the case of the 'Physical Activity Propensity' latent variable, only age is found to be significant in representing or capturing this trait of an individual. It was found that individuals who had the highest propensity to be physically active are in the 12-19 years age group (the youngest age group), which then decreased in magnitude as the individual's age increases. Telama and Yang (2000) found that after age 12, the frequency of participating in at least 30 minutes of physical activity declined continually up to age 27. This trend may be consistent with the notion that individuals may find it difficult to allocate time for leisure activities due to work commitments, responsibilities, and/or maintaining and caring for their family. It is also found that the older generations used exercise to enhance their household chores, were not keen on the idea of gym memberships, and not exposed to physical activity throughout their life (Vertinsky, 1994). This along with their tendency to have more health-related problems, could be preventing them from having a higher propensity for physical activity. In addition, those in the 12-19 year age group may participate in organized sports activities inside or outside of their educational setting. Such activities are likely to diminish with advancing age. Furthermore, it is likely that the attitude of the parents influences the physical activity propensity of the children, while older age groups are independent of such influence.

With respect to the "Health Conscious Attitude" latent variable, the results indicate that the higher the education status of an individual, the higher their level of health consciousness. This may be reflective of educated individuals having better knowledge and greater awareness of the ingredients necessary to lead a healthy life, and having the resources necessary to implement these ingredients (Kenkel, 1991). The gender variable also contributes to "health consciousness", with men being less health-conscious than women. Wardle et al (2004) report similar findings stating that women attach a greater importance to healthy eating and are more likely to be conscious of their nutritional intake than their male counterparts.

The correlation coefficient between the "Health Conscious Attitude" and "Physical Activity Propensity" latent constructs is statistically significant and positive at a value 0.095, indicating that individuals who are more health-conscious are also more inclined to participate in physical activity than their less health-conscious peers.

4.3 Latent Variable Measurement Equation Results (Non-Nominal Outcomes)

Table 3 provides results for the latent variable measurement equation components, which consists of twelve indicator variables (four continuous, three group, four count, and one ordinal). The set of twelve indicator variables includes two health outcome variables and ten physical activity and activity/time-use variables. The results of each component are discussed in this section.

The first component corresponds to the natural logarithm of walk duration, one of the continuous variables. It is found that those who have a normal weight body mass index (BMI) tend to walk for less duration compared to those with higher (overweight and obese) BMI values. Evidence in the US supports this as it has been reported that the highest percentage of people walking are from the overweight and obese BMI categories (Simpson et al, 2003). Those with allergies and asthma are likely to devote less time to walking, presumably because walking outdoors aggravates these conditions creating significant discomfort to the individual (Foty et al, 2013). Walk duration is also seen to vary by race and gender with Hispanics being found to walk longer, while Caucasians walk less. The "Health Conscious" latent variable is found to have a positive impact on walk duration. This latent construct also has a positive influence on time spent bicycling (Bike Duration) and in fact, is the only explanatory factor that appears significant in the model specification. Clearly, health conscious individuals allocate more time to walking and bicycling. The error variances of these two continuous components is fixed to 1.0, because they were very close to 1.0 in most specifications, and we noticed that fixing these led to accelerated convergence.

The time allocated to moderate physical activities is significantly lower for those who fall within the normal or overweight BMI categories. It appears that obese individuals allocate more time to moderate physical activities, presumably in an attempt to improve their health. It is found that Caucasian males have a positive impact on activity duration. Those with a normal or overweight BMI spend less time on vigorous activities compared to their obese BMI counterparts, who may be attempting to rectify their health through physical activity engagement. Those with allergies spend less time pursuing both moderate and vigorous physical activities. Caucasians, African-American females, and Hispanic females were all found to devote less time to vigorous activities.

Only African-American males were found to pursue longer periods of vigorous physical activities.

The self-reported health rating results suggest that those who are obese, diabetic, or have asthma are likely to report a lower state of health, while those who participate in at least 30 minutes of physical activity on a regular basis are likely to report a higher level of health well-being. These findings are consistent with those reported by Okosun et al (2001) who found an inverse association between the proportion of individuals reporting excellent health status and the proportion of those who were obese. As expected, health conscious individuals are more likely to report a higher health rating.

The first grouped variable, blood pressure, is influenced by a number of indicators and socio-demographic variables. Blood pressure is higher for individuals who are overweight, obese, or diabetic. The number of hours watching TV (sedentary activity) is associated with higher levels of blood pressure, a finding consistent with that reported by Sugiyama et al (2007). Along similar lines, the results indicate that participating in any type of physical activity for 30 minutes or more per day will lower blood pressure. Finally, the results indicate that females in general have lower blood pressure than their male counterparts (see Reckelhoff, 2001, for similar findings). It is interesting to note that neither of the latent constructs significantly affects blood pressure; it is likely that blood pressure is more of a medical outcome as opposed to an activity/time-use outcome. While the latter is clearly impacted by latent constructs (representing lifestyle), the former is affected by the activity/time-use patterns rather than the latent constructs per se.

The number of hours spent watching television is affected by the "Physical Activity Propensity" latent variable. Those who have a higher propensity to engage in physical activities devote less time to watching television, a finding consistent with expectations. On the other hand, the time allocated to activities on the computer is affected by the education status of the individual rather than any health or physical activity related variables. Consistent with findings in the American Time Use Survey (Bureau of Labor Statistics 2013), individuals with a higher level of education spend more time on the computer.

The number of walking episodes (a count variable) is affected by several variables. As expected, those who are health conscious (latent construct) are inclined to undertake a greater number of walking episodes. Those who are normal weight report fewer walking episodes while those who are overweight report a higher number of walking episodes. It appears that the overweight individuals are attempting to shed some of the weight through walking activities (Simpson et al, 2003). Hispanics are likely to pursue a greater number of walking episodes; this may, at least in part, be due to the lower levels of auto ownership in Hispanic households resulting in these households walking more than other groups.

For the number of biking episodes (a count variable), it is found that gender plays an important part as does age. Hispanic and African-American females pursue fewer biking episodes than males of any ethnicity, a finding consistent with that reported in past studies that women, especially minorities, exercise less frequently (see Eyler et al, 1999, Garrard et al, 2008). The 12-19 years age group reports a higher number of biking episodes, consistent with the notion that teenagers (with lower levels of auto ownership) are likely to use the bicycle to meet their mobility needs. Winters et al (2007) found that individuals in the 12 to 19 year age group were three times more likely to bicycle than their older counterparts and that bicycling rates decreased steadily with age. Health conscious individuals are likely to pursue a greater number of bicycling episodes.

Those in the normal weight and overweight categories are likely to pursue fewer moderate physical activity episodes in comparison to their obese counterparts. Individuals in the obese category may be attempting to pursue a higher number of such episodes in an attempt to improve their health. Individuals with allergies report undertaking fewer moderate activity episodes, consistent with the notion that such condition limits the ability of individuals to undertake physical activities particularly outdoors. It is interesting to note that individuals with asthma were found to pursue a higher number of moderate episodes. Females, regardless of race, were found to report a larger number of vigorous physical activity episodes, relative to their male counterparts. The relationship between gender and moderate physical activity episode engagement is less clear. For both moderate and vigorous physical activity engagement, the physical activity propensity (latent construct) is found to be an important and significant predictor. The significant dispersion parameters on the count dependent variables show that there is considerable heterogeneity in the population with respect to the frequencies of these variables (variance is greater than mean).

4.4 Latent Variable Measurement Equation Results (Nominal Outcome)

Table 4 displays estimation results for the nominal outcome variable, body mass index. The base category is that corresponding to normal weight. The constants suggest that there is a negative baseline propensity associated with being overweight and a positive baseline propensity associated with being obese. In other words, within this sample, individuals are more likely to fall into the obese category relative to other categories all else being equal. The results also indicate that using the computer for more than three hours per day contributes to obesity, a finding consistent with extant literature (Moto et al, 2006). Individuals with asthma or allergies are likely to be obese. Being diabetic is also associated with being overweight and obese, a finding consistent with that reported in previous studies (e.g., Resnick et al*,* 2000). It is interesting to note that higher income individuals are more likely to be overweight. Higher income individuals may be more auto-centric in their mode choice, spend more time at work (to earn income) reducing time available for healthy physical activities, and have larger caloric intakes as they dine out more often than their lower income counterparts.

The table also presents the results of the loading of the latent variables on the nominal body mass index outcome, with the base category being obese BMI. The two latent variables, Health Conscious Attitude and Physical Activity Propensity, were statistically significant for the overweight BMI group, albeit with opposite signs. Health conscious individuals show a reduced propensity to be overweight, as expected; however, it appears that they exhibit an equal proclivity towards being normal weight or obese. It is possible that health consciousness is associated with these two categories for different reasons. Health conscious individuals maintain good health and are of normal weight, or at the other extreme, obese individuals are health conscious as they attempt to improve their health condition. Physical activity propensity is positively associated with being normal weight or overweight, as opposed to obese, a finding that is intuitive and reasonable. Physical activity increases daily energy expenditures which can assist in weight loss and maintaining a healthy weight (Hills et al, 2011).

4.4.1 Variance-Covariance Parameters

The estimated variance-covariance structure for BMI is presented in Table 4. The estimated error matrix for BMI is significantly different from an IID structure. The variance term for the obese (1.369) is relatively larger than that of overweight and normal weight. There is also a significant positive correlation (implied correlation of 1.00) between overweight and obese, suggesting the presence of common unobserved factors that affect the likelihood of an individual being overweight and obese.

4.5 MEASURES OF FIT

The performance of the GHDM model structure used here may be compared to that of model which does not consider latent constructs (NL-GHDM). The model which does not consider latent construct does not account for dependencies across different modeling dimensions (non-nominal and nominal outcome variables). The composite log-likelihood value for the GHDM model (with 117 parameters) is -446921.83, while the corresponding value for the NL-GHDM model (with 111 parameters) is -491066.94. The two models (GHDM and NL-GHDM) may be compared using the adjusted composite likelihood ratio test (ADCLRT) statistics that is approximately chi-squared distributed (see Bhat, 2011 for a detailed discussion). The ADCLRT statistic value is 504.55, which is larger than the critical chi-square value with 6 degree of freedom at any level of significance. This clearly illustrates the need to consider the dependencies across various modeling dimensions, which can be done efficiently using Bhat's (2014) GHDM model.

CHAPTER 5 CONCLUSIONS

There is increasing interest in drawing connections between activity-travel indicators and public health outcomes. Activity-based models of travel are increasingly providing richer disaggregate information about the types of activities. Many of the indicators related to physical activity participation, sedentary activity participation (such as watching television or sitting at the computer for extended periods), and extent of bicycling and walking are measures that public health professionals would be interested in connecting to health outcomes such as body mass index (BMI), blood pressure, and overall state of health.

However, despite the widespread recognition of the importance of attitudes and lifestyle preferences on activity engagement patterns and mode use, activity-based models fail to include such variables in the model specification. Engagement in physical activities, and the use of bicycle and walk modes, are likely to be influenced by the lifestyle preferences and attitudes of individuals. However, such lifestyle preferences and attitudes are rarely, if ever, measured in surveys rendering it difficult to explicitly include such measures in activity model specifications.

This study constitutes an initial attempt to fill this gap by adopting a GHDM model system in which latent constructs that describe an individual's health consciousness and physical activity propensity are modeled as a function of observed socio-economic and demographic characteristics. The resulting latent constructs, together with socio-economic and demographic variables, are then used to predict a number of activity engagement outcomes (describing frequency and duration of participation in various types of activities – both physically active and sedentary) and health outcomes (such as body mass index, self-reported health well-being, and blood pressure). The entire system of equations is estimated simultaneously through the use of the maximum approximate composite marginal likelihood (MACML) estimation approach that greatly simplifies the evaluation of the likelihood function and brings about computational efficiency in the estimation of simultaneous equations model systems that involve a mixture of dependent variable types.

The findings of the paper show that latent constructs, health consciousness and physical activity propensity, are related to socio-economic and demographic variables. These latent constructs play a significant role in shaping activity-travel and mode use patterns, with those who are more health conscious or inclined towards physically active lifestyles reporting higher levels of physical activity engagement and better health outcomes. Given the significance of the latent variables in explaining activity engagement and mode use, activity-based microsimulation models may be enriched in terms of the model specification through the inclusion of such latent variables that are themselves functions of observed socio-economic and demographic variables collected in travel surveys. There has been a reluctance historically to include attitudinal and lifestyle preference variables in model specifications because such variables are not typically measured in travel surveys, and more importantly, they are difficult to forecast into the future. However, the approach proposed in this paper, where latent variables are functions of observed variables and can be included in models of activity-travel behavior, offers a mechanism by which such latent attitudinal and lifestyle constructs can be included in models of activity-travel demand.

The study is not without its limitations. Due to the nature of the study, the survey data set used for this modeling effort had to include both activity-travel indicators as well as health indicators. The 2005-2006 National Health and Nutrition Examination Survey (NHANES) offered such a data set, but this data set suffered from the drawback that it did not include any built environment, contextual, or network level of service variables – all of which invariably affect activity-travel indicators and possibly health outcomes as well. Furthermore, the data set can include a measure of response bias due to the self-reported state of health and participation in physical activities. Future research and data collection efforts should attempt to include all of the variables of interest so that contextual variables may be accounted for in the model specification.

Appendix A: Methodology Drawn from Bhat (2014)

To estimate the model, note that, under the utility maximization paradigm, $U_{gi_g} - U_{gm_g}$ must be less than zero for all $i_g \neq m_g$ corresponding to the *gth* nominal variable, since the individual chose alternative m_g. Let $u_{gi_g m_g} = U_{gi_g} - U_{gm_g} \, (i_g \neq m_g)$, and stack the latent utility differentials into a vector $\boldsymbol{u_g} = \left[\left(u_{g1m_g}, u_{g2m_g}, \ldots, u_{gI_g m_g} \right)'; i_g \neq m_g \right]$. Also, define $\boldsymbol{u} = \left([\boldsymbol{u_1}]', [\boldsymbol{u_2}]', \ldots, [\boldsymbol{u_G}]' \right)'$. We now need to develop the distribution of the vector $\boldsymbol{yu} = (\vec{\boldsymbol{y}}', \boldsymbol{u}')'$ from that of $\boldsymbol{yU} = [\vec{\boldsymbol{y}}', \boldsymbol{U}']'$. To do so, define a matrix $\mathbf{M}$ of size $[E + \tilde{G}] \times [E + \vec{G}]$. Fill this matrix with values of zero. Then, insert an identity matrix of size *E* into the first *E* rows and *E* columns of the matrix ***M***. Next, consider the rows from $E+1 \text{ to } E+I_1-1$, and columns from $E+1 \text{ to } E+I_1$. These rows and columns correspond to the first nominal variable. Insert an identity matrix of size $(I_1 - 1)$ after supplementing with a column of '-1' values in the column corresponding to the chosen alternative. Next, rows $E+I_1$ through $E+I_1+I_2-2$ and columns $E+I_1+1$ through $E+I_1+I_2$ correspond to the second nominal variable. Again position an identity matrix of size $(I_2 - 1)$ after supplementing with a column of '-1' values in the column corresponding to the chosen alternative for the second nominal variable. Continue this procedure for all *G* nominal variables. With the matrix ***M*** as defined, we can write $\boldsymbol{yu} \sim MVN_{E+\tilde{G}}(\tilde{\boldsymbol{B}}, \tilde{\boldsymbol{\Omega}})$, where $\tilde{\boldsymbol{B}} = \mathbf{M}\boldsymbol{B}$ and $\tilde{\boldsymbol{\Omega}} = \mathbf{M\Omega M'}$. Next, partition the vector $\tilde{\boldsymbol{B}}$

into components that correspond to the mean of the vectors $\boldsymbol{y}$ (for the continuous variables), $\ddot{\boldsymbol{y}} = \left(\left[\tilde{y}^{*}\right]', \left[\vec{y}^{*}\right]', \left[\breve{y}^{*}\right]'\right)'$ $[(N+V+C)\times 1\ vector]$, (for the ordinal, group and count outcomes), and $\boldsymbol{u}$ (for the nominal outcomes), and the matrix $\tilde{\boldsymbol{\Omega}}$ into the corresponding variances and covariances:

$$\tilde{\boldsymbol{B}} = \begin{bmatrix} \tilde{\boldsymbol{B}}_{y} \\ \tilde{\boldsymbol{B}}_{\ddot{y}} \\ \tilde{\boldsymbol{B}}_{\boldsymbol{u}} \end{bmatrix} \ (E+\tilde{G})\times 1 \text{ vector and } \tilde{\boldsymbol{\Omega}} = \begin{bmatrix} \tilde{\boldsymbol{\Omega}}_{y} & \tilde{\boldsymbol{\Omega}}_{y\ddot{y}} & \tilde{\boldsymbol{\Omega}}_{yu} \\ \tilde{\boldsymbol{\Omega}}'_{y\ddot{y}} & \tilde{\boldsymbol{\Omega}}_{\ddot{y}} & \tilde{\boldsymbol{\Omega}}_{yu} \\ \tilde{\boldsymbol{\Omega}}'_{yu} & \tilde{\boldsymbol{\Omega}}'_{\ddot{y}u} & \tilde{\boldsymbol{\Omega}}_{\boldsymbol{u}} \end{bmatrix} (E+\tilde{G})\times(E+\tilde{G}) \text{ matrix} \quad (1)$$

Define $\tilde{\boldsymbol{u}} = \left(\ddot{\boldsymbol{y}}', \boldsymbol{u}'\right)'$, so that $\boldsymbol{yu}=(\boldsymbol{y}',\tilde{\boldsymbol{u}}')'$. Re-partition $\tilde{\boldsymbol{B}}$ and $\tilde{\boldsymbol{\Omega}}$ in a different way such that:

$$\tilde{\boldsymbol{B}} = \begin{bmatrix} \tilde{\boldsymbol{B}}_{y} \\ \tilde{\boldsymbol{B}}_{\tilde{u}} \end{bmatrix}, \text{ where } \tilde{\boldsymbol{B}}_{\tilde{u}} = \begin{bmatrix} \tilde{\boldsymbol{B}}_{\ddot{y}} \\ \tilde{\boldsymbol{B}}_{\boldsymbol{u}} \end{bmatrix} (N+V+C+\tilde{G})\times 1 \text{ vector, and} \quad (2)$$

$$\tilde{\boldsymbol{\Omega}} = \begin{bmatrix} \tilde{\boldsymbol{\Omega}}_{y} & \tilde{\boldsymbol{\Omega}}_{y\tilde{u}} \\ \tilde{\boldsymbol{\Omega}}'_{y\tilde{u}} & \tilde{\boldsymbol{\Omega}}_{\tilde{u}} \end{bmatrix}, \ \tilde{\boldsymbol{\Omega}}_{\tilde{u}} = \begin{bmatrix} \tilde{\boldsymbol{\Omega}}_{\ddot{y}} & \tilde{\boldsymbol{\Omega}}_{\ddot{y}u} \\ \tilde{\boldsymbol{\Omega}}'_{\ddot{y}u} & \tilde{\boldsymbol{\Omega}}_{\boldsymbol{u}} \end{bmatrix} (N+V+C+\tilde{G})\times(N+V+C+\tilde{G}), \text{ and } \tilde{\boldsymbol{\Omega}}'_{y\tilde{u}} = \begin{bmatrix} \tilde{\boldsymbol{\Omega}}_{y\ddot{y}} \\ \tilde{\boldsymbol{\Omega}}_{yu} \end{bmatrix}$$

.

The conditional distribution of $\tilde{\boldsymbol{u}}$, given $\boldsymbol{y}$, is MVN with mean $\bar{\boldsymbol{B}}_{\tilde{u}} = \tilde{\boldsymbol{B}}_{\tilde{u}} + \tilde{\boldsymbol{\Omega}}'_{y\tilde{u}}\,\tilde{\boldsymbol{\Omega}}_{y}^{-1}\left(\boldsymbol{y}-\tilde{\boldsymbol{B}}_{y}\right)$ and variance $\bar{\boldsymbol{\Omega}}_{\tilde{u}} = \tilde{\boldsymbol{\Omega}}_{\tilde{u}} - \tilde{\boldsymbol{\Omega}}'_{y\tilde{u}}\,\tilde{\boldsymbol{\Omega}}_{y}^{-1}\tilde{\boldsymbol{\Omega}}_{y\tilde{u}}$. Next, define threshold vectors as follows:

$\bar{\boldsymbol{\psi}}_{low} = \left[\tilde{\boldsymbol{\psi}}'_{low}, \vec{\boldsymbol{\psi}}'_{low}, \breve{\boldsymbol{\psi}}'_{low}, \left(-\boldsymbol{\infty}_{\tilde{G}}\right)',\right]'$ $([(N+V+C+\tilde{G})\times 1]\ \text{vector})$ and

$\bar{\boldsymbol{\psi}}_{up} = \left[\tilde{\boldsymbol{\psi}}'_{up}, \vec{\boldsymbol{\psi}}'_{up}, \breve{\boldsymbol{\psi}}'_{up}, \left(\boldsymbol{0}_{\tilde{G}}\right)'\right]'$ $([(N+V+C+\tilde{G})\times 1]$ vector), where $-\boldsymbol{\infty}_{\tilde{G}}$ is a $\tilde{G}\times 1$-

column vector of negative infinities, and $\boldsymbol{0}_{\tilde{G}}$ is another $\tilde{G} \times 1$-column vector of zeros. Then the likelihood function may be written as:

$$L(\boldsymbol{\theta}) = f_H(\boldsymbol{y} \,|\, \tilde{\boldsymbol{B}}_y, \tilde{\boldsymbol{\Omega}}_y\,) \times \Pr\left[\overline{\boldsymbol{\psi}}_{low} \leq \tilde{\boldsymbol{u}} \leq \overline{\boldsymbol{\psi}}_{up}\right],$$

$$= f_H(\boldsymbol{y} \,|\, \tilde{\boldsymbol{B}}_y, \tilde{\boldsymbol{\Omega}}_y\,) \times \int_{D_{\tilde{u}}} f_{N+V+C+\tilde{G}}(\tilde{\boldsymbol{u}} \,|\, \overline{\boldsymbol{B}}_{\tilde{\boldsymbol{u}}}, \overline{\boldsymbol{\Omega}}_{\tilde{\boldsymbol{u}}}\,) d\tilde{u}, \tag{3}$$

where the integration domain $D_{\tilde{u}} = \{\tilde{\boldsymbol{u}} : \overline{\boldsymbol{\psi}}_{low} \leq \tilde{\boldsymbol{u}} \leq \overline{\boldsymbol{\psi}}_{up}\}$ is simply the multivariate region of the elements of the $\tilde{\boldsymbol{u}}$ vector determined by the observed ordinal indicator outcomes, and the range $(-\boldsymbol{\infty}_{\tilde{G}}, \boldsymbol{0}_{\tilde{G}})$ for the utility differences is taken with respect to the utility of the observed choice alternative for the nominal outcome. $f_H(\boldsymbol{y} \,|\, \tilde{\boldsymbol{B}}_y, \tilde{\boldsymbol{\Omega}}_y\,)$ is the MVN density function of dimension H with a mean of $\tilde{\boldsymbol{B}}_y$ and a covariance of $\tilde{\boldsymbol{\Omega}}_y$, and evaluated at $\boldsymbol{y}$. The likelihood function for a sample of Q decision-makers is obtained as the product of the individual-level likelihood functions.

The above likelihood function involves the evaluation of an $N+V+C+\tilde{G}$-dimensional rectangular integral for each decision-maker, which can be computationally expensive. Thus, the MACML approach of Bhat (2011) is used.

Consider the following (pairwise) composite marginal likelihood function formed by taking the products (across the N ordinal variables, the C count variables, and G nominal variables) of the joint pairwise probability of the chosen alternatives for a decision-maker, and computed using the analytic approximation of the multivariate normal cumulative distribution (MVNCD) function.

$$L_{CML}(\boldsymbol{\delta}) = f_H(y \mid \tilde{\boldsymbol{B}}_y, \tilde{\boldsymbol{\Omega}}_y) \times \left(\prod_{n=1}^{N-1} \prod_{n'=n+1}^{N} \Pr(j_n = a_n, j_{n'} = a_{n'}) \right) \times \left(\prod_{v=1}^{V-1} \prod_{v'=v+1}^{V} \Pr(j_v = a_v, j_{v'} = a_{v'}) \right) \times \left(\prod_{c=1}^{C-1} \prod_{c'=c+1}^{C} \Pr(k_c = r_c, k_{c'} = r_{c'}) \right) \times$$
$$\left(\prod_{n=1}^{N} \prod_{v=1}^{V} \Pr(j_n = a_n, j_v = a_v) \right) \times \left(\prod_{n=1}^{N} \prod_{c=1}^{C} \Pr(j_n = a_n, k_c = r_c) \right) \times \left(\prod_{v=1}^{V} \prod_{c=1}^{C} \Pr(j_v = a_v, k_c = r_c) \right) \times$$
$$\left(\prod_{n=1}^{N} \prod_{g=1}^{G} \Pr(j_n = a_n, i_g = m_g) \right) \times \left(\prod_{v=1}^{V} \prod_{g=1}^{G} \Pr(j_v = a_v, i_g = m_g) \right) \times \left(\prod_{c=1}^{C} \prod_{g=1}^{G} \Pr(k_c = r_c, i_g = m_g) \right) \times$$
$$\left(\prod_{g=1}^{G-1} \prod_{g'=g+1}^{G} \Pr(i_g = m_g, i_{g'} = m_{g'}) \right). \tag{4}$$

In the above CML approach, the MVNCD function appearing in the CML function is of dimension equal to (1) two for the second component (corresponding to each pair of observed ordinal outcomes), (2) two for the third component (corresponding to each pair of grouped outcomes), (3) two for the fourth component (corresponding to each pair of count outcomes), (4) two for the fifth component (corresponding to each pair of an ordinal outcome and a grouped outcome), (5) two for the sixth component (corresponding to each pair of an ordinal outcome and a count outcome), (6) two for the seventh component (corresponding to each pair of grouped outcome and a count outcome), (7) I_g for the eighth component (corresponding to each pair of an ordinal variable and a nominal variable), (8) I_g for the ninth component (corresponding to a grouped variable and a nominal variable), (9) I_g for the tenth component (corresponding to a count variable and a nominal variable) and (10) $I_g + I_{g'} - 2$ for the eleventh component (corresponding to a pair of nominal outcomes g and g'). The net result is that the pairwise likelihood function now only needs the evaluation of a cumulative normal distribution function of dimension that is utmost equal to the sum of the alternatives minus two associated with the pair of nominal variables with the two highest number of alternatives.

To explicitly write out the CML function in terms of the standard and bivariate standard normal density and cumulative distribution function, define $\boldsymbol{\omega}_{\Delta}$ as the diagonal matrix of standard deviations of matrix $\boldsymbol{\Delta}$, using $\phi_R(.;\boldsymbol{\Delta}^{**})$ for the multivariate standard normal density function of dimension R and correlation matrix $\boldsymbol{\Delta}^*$ ($\boldsymbol{\Delta}^* = \boldsymbol{\omega}_{\Delta}^{-1}\boldsymbol{\Delta}\boldsymbol{\omega}_{\Delta}^{-1}$), and $\Phi_E(.;\boldsymbol{\Delta}^*)$ for the multivariate standard normal cumulative distribution function of dimension E and correlation matrix $\boldsymbol{\Delta}^*$. Define a set of three selection matrices as follows: (1) $\boldsymbol{S}_{rr'}$ is a $2\times(N+V+C+\tilde{G})$ selection matrix with an entry of '1' in the first row and the r^{th} column, another entry of '1' in the second row and the r'^{th} column, and entries of '0' everywhere else $[r,r'\in(1,2,\ldots(N+V+C))]$, (2) $\boldsymbol{D}_{rg}$ is an $I_g\times(N+V+C+\tilde{G})$ selection matrix with an entry of '1' in the first row and the r^{th} column, an identity matrix of size I_g-1 occupying the last I_g-1 rows and the $N+V+C+\left[\sum_{j=1}^{g-1}(I_j-1)+1\right]^{th}$ through $N+V+C+\left[\sum_{j=1}^{g}(I_j-1)\right]^{th}$ columns (with the convention that $\sum_{j=1}^{0}(I_j-1)=0$), and entries of '0' everywhere else, (3) $\boldsymbol{R}_{gg'}$ is a $(I_g+I_{g'}-2)\times(N+V+C+\tilde{G})$ selection matrix with an identity matrix of size (I_g-1)

occupying the first $(I_g - 1)$ rows and the $N+V+C+\left[\sum_{j=1}^{g-1}(I_j-1)+1\right]^{th}$ through $N+V+C+\left[\sum_{j=1}^{g}(I_j-1)\right]^{th}$ columns (with the convention that $\sum_{j=1}^{0}(I_j-1)=0$), and another identity matrix of size $(I_{g'}-1)$ occupying the last $(I_{g'}-1)$ rows and the $N+V+C+\left[\sum_{j=1}^{g'-1}(I_j-1)+1\right]^{th}$ through $N+V+C+\left[\sum_{j=1}^{g'}(I_j-1)\right]^{th}$ columns; all other elements of $\boldsymbol{R}_{gg'}$ take a value of zero. Also, let $\ddot{\boldsymbol{\Omega}}_{rr'} = \boldsymbol{S}_{rr'}\overline{\boldsymbol{\Omega}}_{\tilde{u}}\boldsymbol{S}'_{rr'}$, $\hat{\boldsymbol{\Omega}}_{rg} = \boldsymbol{D}_{rg}\overline{\boldsymbol{\Omega}}_{\tilde{u}}\boldsymbol{D}'_{rg}$, $\vec{\boldsymbol{\Omega}}_{gg'} = \boldsymbol{R}_{gg'}\overline{\boldsymbol{\Omega}}_{\tilde{u}}\boldsymbol{R}'_{gg'}$, $\mu_{r,up} = \dfrac{\lfloor\overline{\boldsymbol{\psi}}_{\boldsymbol{up}}\rfloor_r - [\overline{\boldsymbol{B}}_{\tilde{u}}]_r}{\sqrt{[\overline{\boldsymbol{\Omega}}_{\tilde{u}}]_{rr}}}$, $\mu_{r,low} = \dfrac{[\overline{\boldsymbol{\psi}}_{\boldsymbol{low}}]_r - [\overline{\boldsymbol{B}}_{\tilde{u}}]_r}{\sqrt{[\overline{\boldsymbol{\Omega}}_{\tilde{u}}]_{rr}}}$, $\rho_{rr'} = \dfrac{[\overline{\boldsymbol{\Omega}}_{\tilde{u}}]_{rr'}}{\sqrt{[\overline{\boldsymbol{\Omega}}_{\tilde{u}}]_{rr}[\overline{\boldsymbol{\Omega}}_{\tilde{u}}]_{r'r'}}}$,

where $\lfloor\overline{\boldsymbol{\psi}}_{\boldsymbol{up}}\rfloor_r$ represents the r^{th} element of $\overline{\boldsymbol{\psi}}_{\boldsymbol{up}}$ (and similarly for other vectors), and $[\overline{\boldsymbol{\Omega}}_{\tilde{u}}]_{vv'}$ represents the rr'^{th} element of the matrix $\overline{\boldsymbol{\Omega}}_{\tilde{u}}$. Then,

$$
\begin{aligned}
L_{CML}(\boldsymbol{\delta}) = {} & \left(\prod_{\mathrm{h}=1}^{\mathrm{H}}\boldsymbol{\omega}_{\tilde{\boldsymbol{\Omega}}_y}\right)^{-1}\varphi_H\left(\left[\boldsymbol{\omega}_{\tilde{\boldsymbol{\Omega}}_y}\right]^{-1}\left[\boldsymbol{y}-\tilde{\boldsymbol{B}}_y\right];\tilde{\boldsymbol{\Omega}}_y^*\right)\times \\
& \left(\prod_{r=1}^{N+V+C-1}\prod_{r'=r+1}^{N+V+C}\left[\begin{array}{l}\Phi_2(\mu_{r,up},\mu_{r',up},\rho_{rr'}) - \Phi_2(\mu_{r,up},\mu_{r',low},\rho_{rr'}) \\ -\Phi_2(\mu_{r',low},\mu_{r,up},\rho_{rr'}) + \Phi_2(\mu_{r,low},\mu_{r',low},\rho_{rr'})\end{array}\right]\right)\times \\
& \left(\prod_{r=1}^{N+V+C}\prod_{g=1}^{G}\Phi_{I_g}\left[\boldsymbol{\omega}_{\hat{\boldsymbol{\Omega}}_{rg}}^{-1}\boldsymbol{D}_{rg}\left\{\overline{\boldsymbol{\psi}}_{\boldsymbol{up}}-\overline{\boldsymbol{B}}_{\tilde{u}}\right\};\hat{\boldsymbol{\Omega}}_{vg}^*\right] - \Phi_{I_g}\left[\boldsymbol{\omega}_{\hat{\boldsymbol{\Omega}}_{rg}}^{-1}\boldsymbol{D}_{rg}\left\{\overline{\boldsymbol{\psi}}_{\boldsymbol{low}}-\overline{\boldsymbol{B}}_{\tilde{u}}\right\};\hat{\boldsymbol{\Omega}}_{vg}^*\right]\right)\times \\
& \left(\prod_{g=1}^{G-1}\prod_{g'=1}^{G}\Phi_{I_g+I_{g'}-2}\left[\boldsymbol{\omega}_{\vec{\boldsymbol{\Omega}}_{gg'}}^{-1}\boldsymbol{R}_{gg'}\left\{-\overline{\boldsymbol{B}}_{\tilde{u}}\right\};\vec{\boldsymbol{\Omega}}_{gg'}^*\right]\right).
\end{aligned} \tag{5}
$$

In Equation (24), the first component corresponds to the marginal likelihood of the continuous outcomes, the second component corresponds to the likelihood of pairs of outcomes across all ordinal, grouped and count outcomes (essentially this combines the second to seventh components of Equation (4)), the third component corresponds to the pairwise likelihood for ordinal/grouped/count outcomes and nominal outcomes (this combines the eighth to tenth components of Equation (4)), and the last component corresponds to the pairwise likelihood for the nominal outcomes (this is also the last component of expression (4)).

Appendix B: Tables

TABLE 1. Descriptive Characteristics of the Sample

Demographic Variables	Sample
Gender	
Female	50.64%
Male	49.36%
Age (years)	
12-19	23.45%
20-44	34.61%
45-64	26.76%
65 or older	15.18%
Ethnicity	
Caucasian	46.07%
African-American	25.15%
Mexican American or Other Hispanic	24.40%
Other Race (Included Multi-Racial)	4.38%
Household Size	
One Person	11.21%
Two People	27.85%
Three People	18.92%
Four People	17.15%
Five People	12.89%
Six People	6.59%
Seven People or More	5.38%
Education Status	
Less than 9th Grade	17.78%
9th through 11th Grade	20.19%
High School Graduate or GED	20.80%
Some College or Associates Degree	24.47%
College Graduate or Above	16.74%
Income Level	
Less than $25,000	28.07%
$25,000 - $44,999	23.40%
$45,000 - $64,999	17.25%
$65,000 or More	31.28%
Health Variables	**Sample**
Blood Pressure	
Normal	49.70%
Prehypertension	34.42%
High Blood Pressure Stage 1	11.43%
High Blood Pressure Stage 2 or Crisis	4.45%

TABLE 1. Descriptive Characteristics of the Sample (Continued)

Body Mass Index (BMI)	
Underweight or Normal weight	36.29%
Overweight	30.24%
Obese	33.47%
Covered by Health Insurance	
Yes	78.81%
No	21.19%
Diabetic	
Yes	8.78%
No	91.22%
Have Asthma	
Yes	14.13%
No	85.87%
Have Allergies	
Yes	31.16%
No	68.84%
Physical Activity Variables	**Sample**
Hours watching TV in a day	
None	1.73%
≤ 1 Hour	30.94%
2 Hours	26.49%
3 Hours	17.39%
4 Hours	10.61%
≥5 Hours	12.84%
Hours using Computer in a day	
None	38.90%
≤1 Hour	42.74%
2 Hours	9.24%
3 Hours	3.53%
4 Hours	2.19%
≥5 Hours	3.41%

TABLE 2. Estimation Results of the Latent Variable Structural Equation Model

Variable	Coefficient	t-stat
Physical Activity Propensity (PAP)		
Age (base is 65 and above)		
12 – 19 years (Yes=1, No=0)	1.325	31.473
20 – 44 years (Yes=1, No=0)	0.500	12.964
45 – 64 years (Yes=1, No=0)	0.204	4.954
Health-conscious (HC)		
Education Status (base is 11^{th} grade or less)		
Some college degree (Yes=1, No=0)	0.031	1.937
Bachelor or post-graduate degree (Yes=1, No=0)	0.278	8.122
Gender (base is male)		
Female (Yes=1, No=0)	0.075	2.228
Correlation Between PAP & HC		
Correlation coefficient between PAP & HC latent constructs	0.095	4.913

TABLE 3. Estimation Results for Latent Variable Measurement Equation - Non-nominal Outcomes

Variable	Coefficient (t-stat)
Natural logarithmic of Walk Duration (Continuous variable)	
Constant	-2.587(-38.067)
Body mass index (base is obese)	
Normal-weight (Yes=1, No=0)	-0.256(-7.120)
Overweight (Yes=1, No=0)	3.112(89.730)
Allergy (Yes=1, No=0)	-0.043(-2.621)
Asthma (Yes=1, No=0)	-0.067(-2.959)
Race and Gender combination	
Hispanic male (Yes=1, No=0)	0.297(5.701)
Caucasian male (Yes=1, No=0)	-0.090(-4.054)
Hispanic female (Yes=1, No=0)	0.382(7.665)
Caucasian female (Yes=1, No=0)	-0.257(-10.556)
Latent Variable	
Health-conscious	1.908(45.023)
Variance	1.000(fixed)
Natural logarithmic of Bike Duration (Continuous variable)	
Constant	-2.178(-65.526)
Latent Variable	
Health-conscious	0.069(3.293)
Variance	1.000(fixed)
Natural logarithmic of Moderate Activity Duration (Continuous variable)	
Constant	-0.867(-16.808)
Body mass index (base is obese)	
Normal-weight (Yes=1, No=0)	-2.390(-47.758)
Overweight (Yes=1, No=0)	-1.356(-33.125)
Allergy (Yes=1, No=0)	-0.107(-3.505)
Race and Gender combination	
Caucasian male (Yes=1, No=0)	0.158(5.887)
Latent Variable	
Physical activity propensity	1.175(50.706)
Variance	0.900(43.751)

Variable	Coefficient (t-stat)
Natural logarithmic of Vigorous Activity Duration (Continuous variable)	
Constant	-0.632(-11.673)
Body mass index (base is obese)	
Normal-weight or Overweight (Yes=1, No=0)	-2.059(-73.844)
Allergy (Yes=1, No=0)	-0.117(-5.812)
Race and Gender combination	
Caucasian male (Yes=1, No=0)	-0.179(-7.018)
African-American male (Yes=1, No=0)	0.130(5.057)
Hispanic female (Yes=1, No=0)	-0.179(-4.636)
Caucasian female (Yes=1, No=0)	-0.290(-11.733)
African-American female (Yes=1, No=0)	-0.247(-6.489)
Latent Variable	
Physical activity propensity	1.353(52.063)
Variance	0.572(46.234)
Health Rating (Ordinal Variable)	
Constant	2.144(125.38)
Body mass index (base is normal)	
Overweight or Obese (Yes=1, No=0)	-0.425(-35.984)
Diabetic (Yes=1, No=0)	-0.669(-40.292)
Asthma (Yes=1,No=0)	-0.205(-15.081)
Participates in any type of physical activity for 30 min or more per day in average	0.339(30.057)
Latent Variable	
Health-conscious	0.175(16.816)
Threshold	
Threshold 1 (fair and good)	1.179(90.075)
Threshold 2 (good and very good)	2.329(164.476)
Threshold 3 (very good and excellent)	3.371(220.823)
Blood Pressure (Group Variable)	
Constant	0.092(11.436)
Body mass index (base is normal weight)	
Overweight or Obese (Yes=1, No=0)	0.067(32.090)
Diabetic (Yes=1, No=0)	0.089(28.338)
TV-hours (base: do not watch television)	
Watches television for 0 – 3 hours per day	0.044(5.619)
Watches television for 4 and more hours per	0.069(8.658)
Female (Yes=1, No=0)	-0.043(-21.781)
Participates in any type of physical activity for 30 min or more per day in average	-0.034(-14.696)
Variance	0.024(79.347)

TABLE 3. Estimation Results for Latent Variable Measurement Equation - Non-nominal Outcomes (Continued)

Variable	Coefficient (t-stat)
TV Hours (Group Variable)	
Constant	0.371(46.818)
Latent Variable	
Physical activity propensity	-0.131(-22.538)
Variance	0.974(95.478)
Computer Hours (Group Variable)	
Constant	-1.522(-110.033)
Education status (base is high school or less)	
Some college degree (Yes=1, No=0)	0.348(20.600)
Bachelor or post-graduate degree (Yes=1,	0.497(21.947)
Variance	1.471(85.017)
Number of Walking Episodes (Count	
Constant	-1.312(-12.216)
Body mass index (base is obese)	
Normal-weight (Yes=1, No=0)	-0.211(-4.613)
Overweight (Yes=1, No=0)	2.845(18.228)
Race and Gender combination	
Hispanic male (Yes=1, No=0)	0.225(3.042)
Hispanic female (Yes=1, No=0)	0.192(2.432)
Latent Variable	
Health-conscious	2.471(37.212)
Dispersion parameter	0.103(9.718)
Number of Biking Episodes (Count variable)	
Constant	-0.234(-2.704)
Race and Gender combination	
Hispanic male (Yes=1, No=0)	0.561(3.452)
Caucasian male (Yes=1, No=0)	0.406(2.992)
African-American male (Yes=1, No=0)	0.508(3.253)
Hispanic female (Yes=1, No=0)	-0.944(-7.579)
African-American female (Yes=1, No=0)	-0.354(-2.721)
Age (base is greater than 19 years old)	
12 – 19 years (Yes=1, No=0)	0.373(3.579)
Latent Variable	
Health-conscious	0.116(9.825)
Dispersion parameter	0.032(39.212)
Number of Moderate Activity Episodes (Count	
Constant	1.451(15.020)
Body mass index (base is obese)	
Normal-weight (Yes=1, No=0)	-1.675(-20.621)
Overweight (Yes=1, No=0)	-1.350(-16.876)
Allergy (Yes=1, No=0)	-0.064(-1.988)
Asthma (Yes=1,No=0)	0.094(2.343)
Race and Gender combination	
Hispanic female (Yes=1, No=0)	0.134(3.247)
Caucasian female (Yes=1, No=0)	0.102(2.9112)
African-American female (Yes=1, No=0)	0.218(4.713)
Latent Variable	
Physical activity propensity	0.855(61.037)
Dispersion parameter	0.085(23.372)
Number of Vigorous Activity Episodes (Count	
Constant	0.123(1.674)
Body mass index (base is obese)	
Normal-weight (Yes=1, No=0)	-1.816(27.983)
Overweight (Yes=1, No=0)	-1.722(-26.088)
Race and Gender combination	
Caucasian male (Yes=1, No=0)	-0.092(-3.198)
African-American male (Yes=1, No=0)	0.170(5.155)
Hispanic female (Yes=1, No=0)	0.031(2.030)
Caucasian female (Yes=1, No=0)	-0.206(-6.943)
African-American female (Yes=1, No=0)	-0.029(-1.834)
Latent Variable	
Physical activity propensity	1.659(68.432)
Dispersion parameter	0.046(10.230)

TABLE 4. Estimation Results for Latent Variable Measurement Equation – Nominal Outcome

Body Mass Index (Base: Normal weight)	Overweight		Obese	
Variable	**Coefficient**	**t-stat**	**Coefficient**	**t-stat**
Constant	-1.056	-4.180	0.879	4.237
TV-hours (base is < 3 hours per day)				
More than 3 hours per day			-0.275	-3.852
Computer hours (base is < 3 hours per day)				
More than 3 hours per day			0.507	3.953
Allergy (Yes=0, No=0)			0.378	3.618
Asthma (Yes=0, No=0)			0.525	4.090
Income (base is less than 25,000)				
25,000 – 44,999	0.254	2.311		
45,000 and more	0.467	3.990		
Diabetic (Yes=1, No=0)	0.662	3.784	0.296	2.937
Race and Gender combination				
Hispanic female (Yes=1, No=0)	-1.565	-4.407	0.188	1.972
African-American male (Yes=1, No=0)	-0.277	-2.057	0.435	3.470
Hispanic male	-1.007	-3.563		
Effect of Latent Constructs on Nominal Outcome				
Variable	**Normal weight**		**Overweight**	
Physical activity propensity	3.128	4.540	2.734	4.420
Health-conscious			-5.621	-5.867
Error Difference Matrix				
	Overweight		**Obese**	
Overweight	1.00	(fixed)	1.170	3.302
Obese			1.369	2.210

Appendix C: Figures

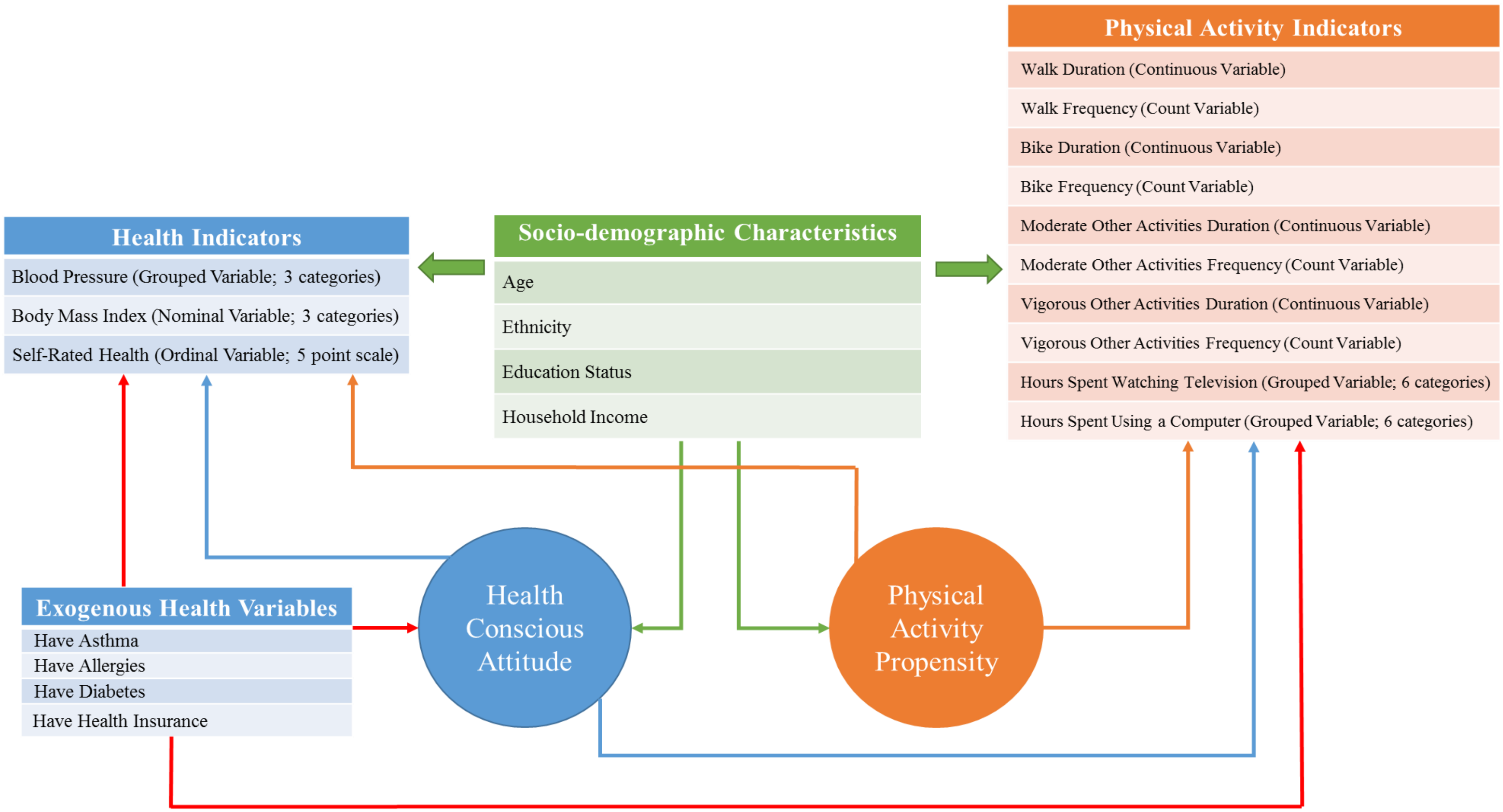

FIGURE 1. Model Framework

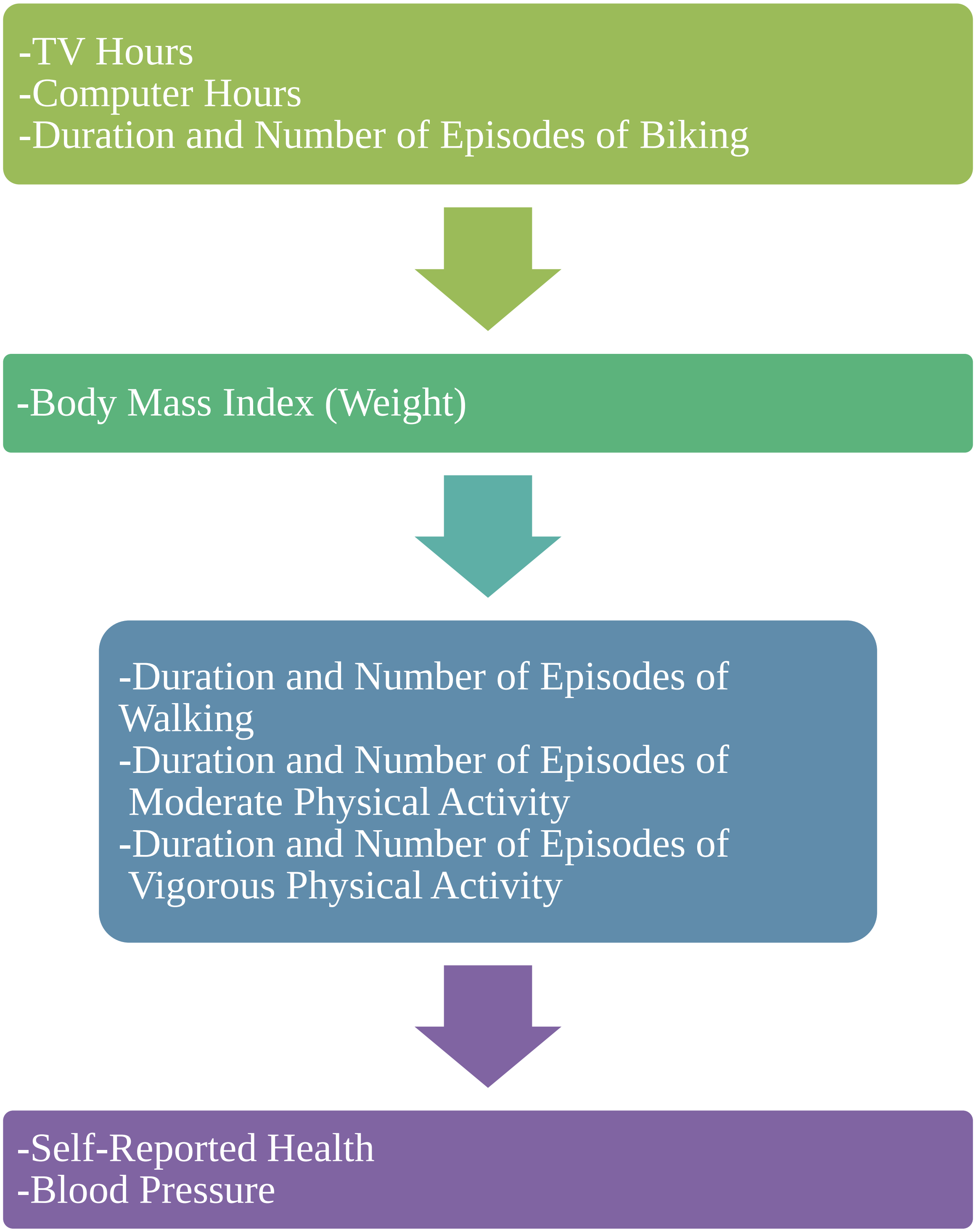

FIGURE 2. Endogenous Effects

References

Arentze, T., and H. Timmermans. Social Networks, Social Interactions, and Activity-Travel Behavior: A Framework for Microsimulation. *Environment and Planning B: Planning and Design*, Vol. 35, 2008, pp. 1012-1027.

Basterfield, L. A. Adamson, J. Frary, K. Parkinson, M. Pearce, J. Reily. Longitudinal Study of Physical Activity and Sedentary Behavior in Children. *Official Journal of the American Academy of Pediatrics*, Vol. 27, No. 24, 2010.

Bauman, A., R. Reis, J. Sallis, J. Wells, R. Loos, B. Martin. Correlates of physical activity: why are some people physically active and others not? *The Lancet,* 2012, pp. 258-271.

Bhat, C.R., S. Astroza, R. Sidharthan, M. Jobair Bin Alam, WH Khushefati. A Joint Count-Continuous Model of Travel Behavior with Selection Based on a Multinomial Probit Residential Density Choice Model. Technical Paper: University of Texas at Austin, 2013.

Bhat, C.R. The maximum approximate composite marginal likelihood (MACML) estimation of multinomial probit-based unordered response choice models. *Transportation Research Part B*, Vol. 45, No. 7, 2011, pp. 923-939.

Bhat, C.R. (2014) A new generalized heterogeneous data model (GHDM) to jointly model mixed types of dependent variables, Technical Paper, Department of Civil, Architectural, and Environmental Engineering, The University of Texas at Austin, available at http://www.caee.utexas.edu/prof/bhat/abstracts/GHMD.pdf.

Bhat, C.R., and Dubey, S.K. A new estimation approach to integrate latent psychological constructs in choice modeling. *Transportation Research Part B*, Vol. 67, 2014, pp. 68-85.

Bhat, C.R., and Sidharthan, R. A simulation evaluation of the maximum approximate composite marginal likelihood (MACML) estimator for mixed multinomial probit models. *Transportation Research Part B*, Vol. 45, No. 7, 2011, pp. 940-953.

Bhat, C.R., Paleti, R., Pendyala, R.M., Lorenzini, K., and Konduri, K.C. Accommodating immigration status and self-selection effects in a joint model of household auto ownership and residential location choice. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2382, No. 1, 2013, pp. 142-150.

Blair, S., Y. Cheng, and S. Holder. Is physical activity or physical fitness more important in defining health benefits? *Medicine & Science in Sports and Exercise*, Vol. 33, No. 6 (supplementary), 2001, S379-S399.

Bureau of Labor Statistics. American Time Use Survey, 2013. http://www.bls.gov/news.release/atus.t11.htm. Accessed May 23, 2014.

Castro, M., Eluru, N., Bhat, C.R., and Pendyala, R.M. (2011). Joint model of participation in nonwork activities and time-of-day choice set formation for workers. *Transportation Research Board*, Vol. 2254, 2011, pp. 140-150.

Castro, M., Paleti, R., and Bhat, C.R. (2012). A latent variable representation of count data models to accommodate spatial and temporal dependence: Application to predicting crash frequency at intersections. *Transportation Research Part B,* Vol 46. No. 1, 2012, pp. 253-272.

Frank, L. D. Land Use and Transportation Interaction Implications on Public Health and Quality of Life. *Journal of Planning Education and Research*, Vol. 20, No. 1, 2000, pp. 6-22.

Frank, L. D., B. Saelens, K. Powell, J. Chapman. Stepping towards causation: Do Built Environments or Neighborhood and Travel Preferences Explain Physical Activity, Driving, and Obesity?. *Social Science and Medicine,* Vol. 65, No. 9, 2007, pp. 1898-1914.

Foty, R., K. Nelligan, T. To, D. Stieb and S. Dell. Children with Asthma are Less Likely to Walk to School. *Pediatric Allergy, Immunology, and Pulmonology*, Vol. 26, No. 2, 2013, pp. 81-87.

Garrard, J., G. Rose and S. Lo. Promoting Transportation Cycling For Women: The Role of Bicycle Infrastructure. *Preventive Medicine*, Vol. 46, No. 1, 2008, pp. 55-59.

Hills, A., L. Andersen, and N. Byrne. Physical Activity and Obesity in Children. *British Journal of Sports Medicine*, Vol. 45, No. 11, 2011, pp. 866-870.

Hu, F., T. Li, G. Colditz, W. Willett and J. Manson. Television Watching and Other Sedentary Behaviors in Relation to Risk of Obesity and Type 2 Diabetes Mellitus in Women. *The Journal of the American Medical Association*, Vol. 289, No. 14, 2003, pp. 1785-1791.

Kenkel, D. S. Health Behavior, Health Knowledge, and Schooling. *Journal of Political Economy*, 1991, pp. 287-305.

Kimm, S., N. Glynn, E. Obarzanek, A. Kriska, S. Daniels, B. Barton, and K. Liu. Relation between the Changes in Physical Activity and Body-Mass Index during Adolescence: A Multicenter Longitudinal Study. *The Lancet*, Vol. 366, no. 9482, 2005, pp. 301-307.

Lubans, D. R., C. Boreham, P. Kelly, C. Foster. The relationship between active travel to school and health-related fitness in children and adolescents: a systematic review. *International Journal of Behavioral Nutrition and Physical Activity,* Vol. 8, No. 5, 2011.

Maddala, G. Limited-Dependent and Qualitative Variables in Econometrics. *Cambridge University Press,* Cambridge, UK, 1983.

Merom, D., H. van der Ploeg, G. Corpuz, A. Bauman. Public Health Perspectives on Household Travel Surveys Active Travel Between 1997 and 2007. *American Journal of Preventive Medicine*. Vol. 39, No. 2, 2010, pp. 113-121.

Mitchell, J.A, R.R. Pate, M.W. Beets, P.R. Nader. Time Spent in Sedentary Behavior and Changes in Childhood BMI: a longitudinal study from ages 9 to 15 years. *International Journal of Obesity,* Vol. 37, 2013, pp. 54-60.

Mokhtarian, P. L., and I. Salomon. How Derived is the Demand for Travel? Some Conceptual and Measurement Considerations. *Transportation Research Part A*, Vol. 35, No. 8, 2001, pp. 695-719.

National Heart, Lung, and Blood Institute. What are the health risks of overweight and obesity?, July 13, 2012. http://www.nhlbi.nih.gov/health/health-topics/topics/obe/risks.html, Accessed April 15, 2014.

Okosun, I., S. Choi, T. Matamoros, and A. Dever. Obesity Is Associated with Reduced Self-Rated General Health Status: Evidence from a Representative Sample of White, Black, and Hispanic Americans. *Preventive Medicine*, Vol. 32, No. 5, 2001, pp. 429-436.

Panter, J. R., A. P. Jones, E. M. F. van Sluijs, S. J. Griffin. Attitudes, social support and environmental perceptions as predictors of active commuting behavior in school children. *Journal of Epidemiology and Community Health*, Vol. 64, No. 1, 2010, pp. 41-48.

Reckelhoff, J. Gender Differences in the Regulation of Blood Pressure. *Hypertension*, Vol. 37, No. 5, 2001, pp. 1199-1208.

Reilly, T., and O'Brien, R.M.. Identification of confirmatory factor analysis models of arbitrary complexity the side-by-side rule. *Sociological Methods and Research.* Vol. 24, No. 4, 1996, pp. 473-491.

Resnick, H., P. Valsania, J. Halter, and X. Lin. Relation of Weight Gain and Weight Loss on Subsequent Diabetes Risk in Overweight Adults. *Journal of Epidemiology and Community Health*, Vol. 54, No. 8, 2000, pp. 596-602.

Simpson, M., M. Serdula, D. Galuska, C. Gillespie, R. Donehoo, C. Macera, and K. Mack. Walking Trends among U.S. adults: The Behavioral Risk Factor Surveillance

System, 1987-2000. *American Journal of Preventive Medicine*, Vol. 25, No. 2, 2003, pp. 95-100.

Stapleton, D.C. Analyzing political participation data with a MIMIC Model. *Sociological Methodology.* 1978, pp.52-74.

Steinman, L., M. Doescher, D. Levinger, C. Perry, L. Carter, A. Eyler, and C. Voorhees. Master Plans for Pedestrian and Bicycle Transportation: Community Characteristics. *Journal of Physical Activity and Health*, Vol. 7, No. 1, 2010, pp. S60-S66.

Sugiyama, T., G. Healy, D. Dunstan, J. Salmon, and N. Owen. Is Television Viewing Time a Marker of a Broader Pattern of Sedentary Behavior? *Annals of Behavioral Medicine*, Vol. 35, No. 2, 2008, pp. 245-250.

Telama, R. and X. Yang. Decline of physical activity from youth to young adulthood in Finland. *Medicine and Science in Sports and Exercise*, Vol. 32, No. 9, 2000, pp. 1617-1622.

University of Maryland Medical Center. *Asthma*. June 2011, http://umm.edu/health/medical/altmed/condition/asthma. Accessed August 1, 2014.

U.S. Census Bureau. *Educational Attainment.* September 2010, https://www.census.gov/hhes/socdemo/education/data/cps/2006/tables.html. Accessed November 3, 2014.

U.S. Department of Health and Human Services. Physical Activity and Health: A Report of the Surgeon General. DIANE Publishing, 1996.

U.S. Department of Health and Human Services. Physical Activity Guidelines for Americans, 2008.

Varin, C. On composite marginal likelihoods. *Advances in Statistical Analysis,* Vol. 92, No. 1, 2008, pp.1-28.

Vertinksky, P. The Eternally Wounded Woman: Doctors, Women and Exercise in the Late Nineteenth Century. Manchester: *Manchester University Press*, 1994, pp. 1-279.

Wardle, J., A. Haase, A. Steptoe, M. Nillapun, K. Jonwutiwes, and F. Bellisie. Gender Differences in Food Choice: The Contribution of Health Beliefs and Dieting. *Annals of Behavioral Medicine*, Vol. 27, No. 2, 2004, pp. 107-116.

Winters, M., M. Friesen, M. Koehoorn, and K. Teschke. Utilitarian Bicycling: A Multilevel Analysis of Climate and Personal Influences. *American Journal of Preventive Medicine*, Vol. 32, No. 1, 2007, pp. 52-58.

Zhao, Y., and Joe, H. Composite likelihood estimation in multivariate data analysis. *Canadian Journal of Statistics*, Vol. 33, No. 3, 2005, pp. 335-356.

Zhang, T J., D. Seo, L. Kolbe, S. Middlestadt, W. Zhao. Associated Trends in Sedentary Behavior and BMI Among Chinese School Children and Adolescents in Seven Diverse Chinese Provinces. *International Journal of Behavioral Medicine,* Vol. 19, 2012, pp. 342-350.